The Effects of Different Error Correction Conditions on Learner-initiated Noticing

María Elena Solares Altamirano

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy at the University of Lancaster.

Department of Linguistics and English Language

Lancaster University

December 2015

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. No parts of this thesis have been submitted for any other qualification.

The thesis does not exceed the regulation length, including footnotes, references and appendices but excluding the bibliography.

This research was supported by the PASPA (Programa de Apoyo a la Superación del Personal Académico) Programme, Dirección General de Asuntos del Personal Académico, Universidad Nacional Autónoma de México.

The William Ritchie and Friend's Programme Travel Fund is also acknowledged for its financial support to present the results of this research at the Symposium of Second Language Writing (Arizona, October 2016).

Acknowledgements

Embarking on PhD study is a tough decision for many people. Doing it because you want to better understand the job you enjoy is different. Foreign languages have long been an important part of my life. The road from language learner to teacher trainer, course designer and eventually researcher has been full of discoveries. The L2 learning process and the EFL (English as a Foreign Language) classroom have always been central to this journey. Language teaching and learning, communication and pedagogy merge to build the richness of FL classrooms. Researchers believe they have knowledge of and expertise in how languages are learned. Quite frequently, however, these are too distant from everyday classroom realities. Teachers, on the other hand, have much intuitively gained knowledge and experience, and many practical skills. But they frequently lack the theoretical and methodological basis to justify their intuition or turn it into knowledge. During my teaching career, a distance between teachers and researchers has always existed. Hence, it was my desire to develop myself professionally and narrow this gap, at least for myself. I thank the people, family and friends who have supported me in achieving this goal.

First, pursuing postgraduate study at a renowned institution like Lancaster University was only possible thanks to the economic support I received from the Dirección General de Asuntos del Personal Académico through its PASPA (Programa de Apoyo a la Superación del Personal Académico) programme. I thank those who positively evaluated my initial grant application and yearly reports. Studying and working at UNAM has been one of the best experiences of my life. To the people who run this university I owe much of what I am and do today. It is them whom I want to thank: my aforementioned sponsors DGAPA (PASPA programme) but especially the

Foreign Language Teaching Centre (CELE) where I have worked for the last 22 years. Colleagues, teachers and students who participated in my research deserve special gratitude: Nicola Geeda Mathews, Barbara Byer and Shelley Chapman, for their detailed work in reformulating students´ essays in the pilot and main studies; my teacher colleagues Haydee Venosa Figueroa, Margarita Ortega Romero and Oscar García Benavides for the endless hours of error correction and discussion during reliability analyses; Alejandra Plaza, Ana Maria Salazar, Jovana Vivanco and Lucia Mena, teacher colleagues who warmly received me in their classrooms and sacrificed various teaching sessions to help me collect my data; the anonymous 45 committed learners who eagerly and responsibly attended each of the writing sessions and shared their learning processes with me. My research would have not been possible without you. CELE students and teachers will benefit from your participation in this project. My thanks extend to all the CELE community: administrators, Dra. Aline Signoret and Lic. Bertha López Escudero, for their faith in me and constantly praising my work. I thank you both for the economic support I received from the Language Centre you represent, and for attending every residential period during this distance PhD programme. Librarians (especially Miguel Angel Ramirez), computer engineers (especially David Morales) and administrative staff, you all contributed to this academic enterprise.

Secondly, but first in my heart, I thank Milton Medeiros de Oliveira, my 'Gatinho', my love, my husband, my support and my sunshine. Your help in statistical analysis was essential, but your love, belief in me, praise, advice, patience and company, and the feelings of fulfilment and security you gave and give me every day, were even more important in this endeavour. You are and will always be my inspiration. Gerard Hearne, my personal proof-reader and friend, I was fortunate to

have you on this journey. We did it! Dr Patrick Rebuschat, my supervisor, you never imposed your views on me and always gave me the confidence and freedom to make my own decisions. Dra. Dulce María Gilbón, my national supervisor and friend, you have always been supportive in academic and administrative matters.

I am also grateful to my nearest and dearest, my family, my mum, Elena, my siblings, Roberto, Fernando, Francisco and Blanca, for your patience, your company, your trust and constant encouragement. You comforted me whenever I needed it. Dad, this is for you too. My nephews and nieces who turned my hardest days into joy and laughs, your love and hugs gave me the energy and will to move on. Finally, one of the greatest gains from any learning period at university is undoubtedly the friends we make and, in my case, the previous friendships I renewed. Friends who share my goals, my dreams, my frustrations and experiences are invaluable. Stef Ecchecevò, Joan Oakley, Winnie Hoo, Hilda Hidalgo, Shaun Austin, Ana Preciado, José Luis Moreno, Pamela Olmos and Anne Bruehler, you made the lonely life of a PhD student fun. Maritza Vázquez, Maria Aurora Palma, Paty Cánovas, José Martín Gasca and Maite Mallén, thank you for always being there for me, for listening to my complaints, moments of disillusion and accomplishments.

To all of you, my greatest appreciation!

Abstract

This research aimed to design a suitable pedagogic study on the effects of different error correction conditions (ECCs) on learner-initiated noticing. With this in mind, I sought to tackle comprehensive (eventually replaced by semi-comprehensive) error correction (EC) in an open, uncontrolled and learner-centred writing task. Despite numerous studies on written corrective feedback (WCF), little research has paid attention to what learners 'notice' while writing. Storch and Wigglesworth (2010) explain that noticing and the "processing of feedback [are] … less … researched … because it is difficult to access such learner-internal cognitive processes" (p. 305). Meanwhile Santos, López-Serrano and Manchón (2010 p. 132) indicate that the self-initiated character of writing problems turns writing into the perfect setting to study self-initiated noticing and focus-on-form processes. Framed within the 'noticing' debate, the 'language learning potential of writing' and the 'writing-to-learn' and 'feedback-for-acquisition' dimensions, this quasi-experimental study investigates how different ECCs influence the error types that learners 'attend to'. The four-stage (composing/ error correction-noticing/ rewriting/ new writing) design involved 60 EFL undergraduates in three semi-comprehensive ECCs (Direct EC, Reformulation and Self-correction). Opinion essays, noticing sheets and a questionnaire constituted the data to analyse. The quantitative results support Truscott's claim for the little value placed on '[semi-] comprehensive' EC in grammar accuracy improvement. Qualitative analyses showed that the ECCs tested had different effects on learners' noticing. Delayed self-correction elicited the most attention to form and is suggested as a more refined way to 'notice the hole'; direct EC led to retention, reformulation resulted in the most engaging ECC.

Table of Contents

Abbreviations

L2: Second language

SLA: Second language acquisition

FL: Foreign language

L1: Mother language

EFL: English as a foreign language

ELT: English language teaching

WCF: Written corrective feedback

CF: Corrective feedback

ECCs: Error correction condition(s)

EC: Error correction

LLPW: Language learning potential of writing

LW: Learning-to-write

WLC: Writing-to-learn content

WLL: Writing-to-learn language

NH: Noticing hypothesis

NS(s): Noticing sheet(s)

DIR/ DIR-G: Direct (group, error correction condition or WCF)

REF/ REF-G: Reformulation (group, error correction condition or WCF)

SELF/ SELF-G: Self (group, error correction condition or WCF)

FN: Features noticed (Hanaoka, 2007)

PFN: Problematic features noticed (Hanaoka, 2007)

RQ(s): Research Question(s)

Q: Question

GR: Grammatical

N-GR: Non-grammatical

UDA: Unnecessary definite article

SECTION 1: CONTEXT, THEORY AND RESEARCH

Chapter 1: Introduction

1.1 Context of the study

I begin with some introductory issues. Chapter 1 is divided into four parts. Part 1 describes the educational and institutional context in which this study was conducted: The English as a foreign language (EFL) context of Mexico and the role of English at the National Autonomous University of Mexico (Universidad Nacional Autónoma de México, UNAM). The pedagogic context of Centro de Enseñanza de Lenguas Extranjeras (CELE) is also described: target students, language courses, syllabus and teaching staff. Part 2 presents common features of EFL classrooms concerning the teaching of writing. Part 3 introduces the topic of the study: WCF in foreign language (FL) writing. Part 4 offers a general outline of this thesis.

1.1.1 Foreign Language Teaching Centre

The role of English as an international language, its role as the lingua franca of science and technology, and the geographical location of Mexico in relation to the United States make the teaching of English as a foreign language an important component of the Mexican language curriculum. Mexico's educational system comprises state (education is free) and private (education is financed by individuals) systems. Educational policies are defined by the Secretaría de Educación Pública (SEP). As required by SEP policies, English has been taught as a compulsory subject after the sixth year of primary education for many years; recently, new education policies have been promoted to make English compulsory in primary education too. However, the small number of pedagogically trained teachers and the budget limitations of state education make the teaching of English quite arduous work for most Mexican teachers and learners, especially those in the state sector. Unless they

come from private schools, or have studied at private language institutions, most university undergraduates arrive at university with low English proficiency.

The National Autonomous University of Mexico (UNAM) is the biggest and most important university in Mexico and Ibero-America, with 115 majors, 41 postgraduate programmes and 342,542 registered students in 2014–2015.[1] UNAM is one of the federal state universities in Mexico. It comprises 15 faculties, four schools, 33 research institutes and 14 centres. As an autonomous university, UNAM has the freedom to define its own curriculum and administer its own budget without interference from the government. UNAM's mission is to educate, research and promote culture. At UNAM and university level in general, advanced knowledge of English and some other foreign languages is compulsory. High English proficiency is required, especially in science and technology. Through its foreign language teaching centre (CELE), UNAM runs foreign languages courses for students to help them meet the academic demands in their particular fields. CELE's evaluation and certification office is in charge of verifying that students have the language skills required by different majors within UNAM. With other institutions, both inside and outside UNAM, CELE shares the responsibility of educating and developing foreign language teachers. Through its Department of Applied Linguistics, CELE carries out research on the teaching and learning of foreign languages, on assessment and on Applied Linguistics in general to succeed in its mission.

Sixteen foreign languages (Arabic, Catalan, Chinese, English, French, German, Greek, Hebrew, Italian, Japanese, Korean, Portuguese, Romanian, Russian, Swedish, Vasco) and one indigenous Mexican language (Náhuatl) are taught at CELE

[1] Data from: UNAM, portal de estadística universitaria http://www.estadistica.unam.mx/numeralia/

to an average of 8,386 students each term.[2] Of that figure, 2, 768 are English students only, followed by the French Department with 1,717 students. Several international exams are currently administered at CELE: Institutional TOEFL (Test of English as a Foreign Language) for English; DELF and DALF (Diplomas of French Language Proficiency) for French; Test-DaF, TestAs and onDaF (German Proficiency Tests) and Österreichische Sprachdiplom Dutch (Austrian Proficiency Test) for German; DUPLE and DAPLE (Diplomas Avançados Português Língua Estrangeira) for Portuguese; the HSK (Hanyu Shaopin Kaoshi) Chinese Proficiency Test for Chinese; the CELI (Certificato de Conoscenza della Lingua Italiana) for Italian.

CELE is composed of various academic departments that contribute to its academic strength. The Department of Applied Linguistics has around 52 specialists in Applied Linguistics. The Evaluation and Certification Department is in charge of certifying UNAM students´ language proficiency. The Teacher Training Department runs one of the most recognized foreign language teacher training programmes in Mexico and an online teacher development diploma course. The Distance Education Department is in charge of promoting the design and development of online courses and e-materials and strengthening research in this area. The Postgraduate Department is a joint initiative with the Schools of Philosophy and Modern Literature and the Philology Institute at UNAM to offer an MA in Applied Linguistics, a PhD in Linguistics and Specialization in the Teaching of Spanish as a Foreign Language. The Department of Translation runs a translators training programme. The self-access centre is in charge of promoting and supporting students´ autonomous learning of different foreign languages and the Publications Department is in charge of publishing

[2] Data from: CELE, Registry Office: February–March 2015.
Boletín Electrónico No. 45, March 2011.
http://cele.unam.mx/boletin/anteriores/boletin_045/html/numeralia/numeralia.html

the centre´s periodic journal 'Estudios en Lingüística Aplicada' (ELA) (Studies in Applied Linguistics), language textbooks and other teaching materials produced at CELE.

1.1.2 CELE students

CELE English students come from different study areas: Social Sciences, Humanities, Engineering, Hard Sciences, Art and Philosophy. Their age range is mainly 19–25 years old, with older academic and administrative members of staff who may also join language courses. According to a 2006 survey, students in the English department study English to satisfy an academic requirement, to study abroad or to improve their job opportunities. However, as neither credits are gained nor requirements met by studying at CELE, English becomes an extracurricular subject and students attend courses on a voluntary basis. A population with different school backgrounds within the multicultural language atmosphere and the fee-free courses make studying at CELE an academically and culturally rich experience that attracts UNAM students not only because of their interest in learning a foreign language, but also for what it means to be part of a community with such a range of academic backgrounds and great cultural richness.

1.1.3 English courses at CELE

Given the large demand for English courses at CELE and the little time (no more than three years) most students have to complete their studies, CELE only offers English courses starting at the A2 level of the Common European Framework of Reference for Languages (CEFR, 2001). According to Vandergrift (n.d.), this level roughly corresponds to the intermediate-mid or intermediate-high descriptors in the American Council on the Teaching of Foreign Languages Proficiency Guidelines

(ACTFL, 1999). Thus language departments in other schools, faculties and colleges at UNAM are responsible for providing basic English instruction. For students registered at UNAM entering the second year of their majors, it is a requirement to study English at CELE and to have at least A2 level English proficiency. The centre´s placement exam acts as a filter.

Graduates, postgraduates and members of the academic and administrative staff are also eligible to study at this language centre. By offering five general English courses (3rd to 7th levels), which are usually covered in an uninterrupted three-year period, students are expected to reach B2 level. Third to seventh English levels are general English, four-skill courses that are followed by three advanced courses (choose from speaking, writing, listening, extensive reading and phonetics). The study presented here was done on four 7th level groups. Other English for specific purposes (ESP) courses are available at CELE: reading comprehension, business English, English for science and technology, English literature and preparation courses for TOEFL. English courses usually last for 96 hours spread over 16 weeks.

1.1.4 The centre's syllabus

Attempts to develop the centre´s own syllabus have been a constant goal ever since it was created. At least four formal, long-term projects have been completed, aiming at creating such a syllabus. However, teachers have described the syllabuses resulting from such projects as too general, too abstract or not practical for everyday classes. Therefore, although the centre might have had its own syllabus in some periods, it was soon replaced by a textbook-based one, usually the syllabus of the textbook in turn at CELE. A textbook-bound context has characterised CELE´s English courses for several years, with departmental exams being developed from textbook content. Various textbooks have been used in the language centre: Headway,

Interchange, Skyline, American Inside Out and New American Cutting Edge, among others. The struggle to make such international textbooks meet the specific needs of Mexican university students has always been a challenge for teachers. The newest textbooks on the market, those claiming to be task-based in their orientation or using a lexical approach for instance, are adopted as a way of paying lip service to the latest findings in English language teaching (ELT). The most recent attempt to implement CELE's own syllabus was in 2013. This was combined with the withdrawal of textbooks. The change was a top-down decision that caused much controversy among teachers. Some were delighted to get rid of textbooks and eager to face a new challenge; however, other teachers panicked at the idea of not having a textbook to rely on. For the last two years, teachers in the English Department have been getting used to the new syllabus and evaluation procedures. It is in this transition context that this research was conducted.

1.1.5 Teaching staff

CELE has approximately 70 teaching staff, most of them of holding a CELE teacher training diploma equivalent to 750 hours of instruction. The age and academic profile of English teachers vary widely, from people with a long academic background and teaching experience to less academic or less experienced teachers. Some of them work for different institutions and only come to CELE for two or three classes each week. The teaching methods at CELE vary. Some teachers may still use traditional form-focused PPP (presentation-practice-production) methods, though more use communicative, project-based and task-based language teaching with plenty of input and incidental learning.

In parallel to the adoption of the new syllabus and the withdrawal of textbooks, CELE has had to respond to technological developments. Foreign language

courses are rapidly moving to online or blended environments. In the last few years, the creation of such environments for language teaching has become a priority at CELE. Online environments in particular are seen as one way of reaching larger number of students with fewer material resources. This has proved true for some teacher development, reading comprehension, extensive reading and writing for publication courses created and implemented at CELE over the last decade. Emphasis has also been put on the incorporation of information and communication technologies (ICTs) into the EFL classroom by developing teachers´ digital literacy and their use of different digital tools. Follow-up on how teachers integrate ICTs into language courses usually takes place informally, through department workshops, meetings and teacher development courses.

1.2 Writing in the EFL classroom

Writing in a foreign or second language (FL or L2) is an academic and professional requirement for university students and professionals, not only in Mexico but in many other countries too. L2 writing is also one of the greatest challenges for language learners. Evans, Hartshorn, McCollum & Wolfersberger (2010), Giraldo de Londoño and Perry (2008), Lee (2011) and Zamel (1985) identify characteristics of L2 classrooms that are relevant for writing research. Here I consider those that are common to foreign language, 'communicative', four-ability classrooms, like the one contextualizing this study. Despite the 'prevalence' of communicative language teaching, project-based and task-based learning, and other recent approaches to L2 teaching, teachers' written feedback practices in FL, four-ability, general English classrooms are characterized by a focus on errors, accuracy and correctness. Zamel (1985) states that "searching for and calling attention to error is still the most widely employed procedure for responding to ESL [EFL] writing" (p. 84).

In CELE's EFL classrooms (ESP writing courses excluded), writing is sometimes used as a means to practise grammar rather than as an end in itself. This may be explained by the full-content syllabus that leaves little time to work on the development of writing as a process in four-skill courses. Nevertheless, most CELE teachers manage to combine grammatical accuracy with other writing skills (text organization, cohesion of ideas or style), confronting learners with the demanding task of attending to all these aspects of writing simultaneously.

The L1 writing skill of CELE EFL students is usually assumed in FL classrooms, i.e. learners are expected to know how to write in their mother tongue. However, learners' ability to write in their L1 is generally underdeveloped. Consequently, learners are usually inexperienced and underprepared to write in a FL.

1.3 Written corrective feedback in FL writing

In the development of L2 writing, feedback or error correction plays a crucial role. Helping learners overcome the errors they make while acquiring the target language constitutes one of the main interests of Second Language Acquisition (SLA) researchers and FL teachers. For researchers, how writing develops, which cognitive processes language learners utilise while writing (Hyland, 2003), what type of learning results from the linguistic processing promoted by writing and whether this linguistic processing leads to learning (Manchón, 2013) are questions without clear answers. For FL teachers, how best to take advantage of one of their most demanding tasks is fundamental.

The topic of WCF gained prominence with Truscott's (1996) questioning of its effectiveness and a call for its eradication. Truscott claims that L2 learning is not a transfer of information as correction implies; it is not realistic to offer WCF at the

point when a learner is ready to acquire a specific structure; it is difficult for one type of WCF to treat different language areas; teachers' competence to give reliable feedback and students' ability and determination to use it effectively are uncertain. WCF may, according to Truscott, maximally result in the development of explicit declarative knowledge rather than implicit procedural knowledge. Over the last two decades, Truscott's claims have prompted endless studies seeking to respond to his assertions and explore the effectiveness of WCF for L2 learning (e.g. Bitchener, 2008; Bruton, 2009; Ferris, 2004; Guénette, 2007 to mention some). The results of research on WCF have, however, been uncertain, partial and contradictory.

Lapkin and Swain (1990) see it as "surprising … that is so little known about the effects of error correction" (p. 655), despite there being so much research. While progress has perhaps been made since Lapkin and Swain' statement, results suggest it has not been substantial. SLA researchers cannot, however, deny that Truscott's claims have translated into much better appreciation of the complexity of writing and of WCF processing. Evans et al. (2010) suggest that the absence of major progress is because researchers have been investigating the wrong problem(s). We do not want to know whether providing WCF is efficient or not, rather we want to know how we can help students to write accurately. Researchers might also be approaching what is a practical problem as a theoretical one. Polio (2012) states that error correction is worth investigating "at a practical level even without reference to specific theories" (p. 376), simply because it is a pedagogical practice prevalent in all learning contexts and consumes a lot of time.

1.4 Thesis outline

This thesis is organised in five sections, each comprising several chapters. Section 1 'Context, Theory and Research' is composed of Chapters 1, 2, 3 and 4.

Chapter 1 describes the educational and institutional context in which this study was conducted. The chapter offers a brief overview of writing in the EFL classroom and WCF in FL writing. The chapter closes with an outline of the thesis. Chapter 2 explains the theoretical framework supporting this study. Chapter 3 conducts a literature review. It starts with a look at Truscott's work, followed by a review of existing experimental studies on WCF, noticing, and noticing in WCF. Chapter 4 explains the gap the study intends to fill, its contribution, its aims and the rationale for its experimental design. The variables of the study, as well as the definitions and operationalisations of constructs are found throughout this chapter.

Section 2 'Pilot Study' contains Chapters 5 and 6. Chapter 5 deals with the methodology (participants, objectives and implementation). Chapter 6 focuses on the implications of the results for the main study.

Section 3 'Towards the Main Study' includes Chapters 7 and 8. Section 3 is an unusual section in the format of experimental studies. It was, however, necessary to include it here to (a) maintain the text coherence and (b) because the outcomes (error categories) from this stage were to be used in the main study. The section considers unexpected methodological problems during the main study regarding unrestrained essay analyses with comprehensive error correction (the written correction technique I originally planned to use). Some problems with reliability analyses obliged me to change to a 'semi-comprehensive' error correction technique. Chapter 7 presents the problems and challenges emerging from the original comprehensive error correction approach. Chapter 8 describes the selected error categories and the rubrics decided upon to shift to 'semi-comprehensive' error correction. To avoid overwhelming the reader, only necessary information is included.

Section 4 ‘Main Study’ comprises Chapters 9, 10 and 11. Chapter 9 presents the research questions, describes the participants and data collection methods and details the treatments and experimental procedure. The last section in Chapter 9 brings together the experimental manipulations and strategies implemented in the design. Chapter 10 explains the data analysis and Chapter 11 presents the results obtained from essays, noticing sheets and an exit questionnaire.

Section 5 ‘Discussion and Conclusion’ is composed of four chapters. Chapter 12 answers the research questions and discusses the findings for each question. Chapter 13 discusses the findings from the questionnaire. Chapter 14 examines some additional findings of the study. Reflections on the theoretical and pedagogical implications of the study are presented in Chapter 15, while Chapter 16 looks at the limitations of the study. Finally, Chapter 17 closes with a general summary and outlines directions for future studies.

Chapter 2: Theoretical Framework

2.1 Cognitive-interactionist framework of SLA

L2 learning in this thesis is approached from the cognitive-interactionist framework of SLA. In this framework, L2 learning entails learning a new knowledge system, and as such, Skehan (1998) explains, it is primarily a cognitive process. Understanding learners' cognitive processes when learning an L2 constitutes one of the central challenges of SLA and cognitive psychology. Researchers in both fields explore learners' mental processes and representations of L2 knowledge in the mind. This framework is concerned with various processes including attention, perception, recognition, comprehension, memory and learning.

The role of noticing in L2 learning (oral and written) has received great support and prompted much investigation (e.g. Adams, 2003; Godfroid, Housen & Boers, 2010; Leow, 1997; Mackey, 2006). However, its function and relationship with awareness make it one of the most controversial cognitive issues in SLA. Since noticing and WCF are at the centre of this study, I deal first with the cognitive framework of SLA that supports my study. I start with the origins and development of the debate around noticing and related terms. Then, I look at the interactionist part of the framework. I define concepts such as input, output, intake and interlanguage. Finally, the writing-to-learn dimension and the learning potential of writing are explained.

2.1.1 Cognitive framework: attention, awareness, noticing

Attention, memory, processing and automaticity are some of the topics that have gained in importance over recent decades. In this study, one of the most controversial cognitive issues is explored: the role of attention and noticing in L2

learning. This has become controversial, not only because of the terminological confusion around noticing and other related terms such as consciousness and awareness, but also because of the disagreements between theorists on the role that consciousness plays in L2 learning. Despite some attempts to differentiate between consciousness and awareness (Schmidt, 1990; Tomlin & Villa, 1994), most theorists use the terms synonymously; thus, they will be used synonymously in this thesis too. Attention was selected as the broadest term that would allow me to explain noticing as an attentional process. Schmidt (1995, p. 18) states that modern psychologists distinguish between attention (as one of the basic mechanisms in an information-processing system) and the correlated subjective experience of noticing, what one attends to and being aware of it. James (in Schmidt, 2001, p. 12) defines attention as "taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought". Schmidt (2001) affirms that attention is essential to understand nearly every aspect of SLA: (1) the development of interlanguage, where attention to input allows for hypothesis formulation and testing; (2) the development of L2 fluency, where automaticity releases attentional resources for other cognitive tasks; (3) the role of individual differences, since learners' attentional capacities differ depending on their aptitudes, motivations and strategies; (4) the way interaction, negotiation for meaning and instruction contribute to L2 learning. Clarification requests and recasts are attention-focusing devices that help acquisition during interaction. Input enhancement and task characteristics are attention-focusing techniques in instructional settings.

Some SLA researchers (Baars, 1988; Van Lier, 1991; Van Patten, 1990, 1994, 1996) have looked at psychologists' studies on attention and incorporated their findings in the SLA field. SLA researchers presume that attention (a) is limited

(McLaughlin, Rossman & McLeod, 1983; Van Patten, 1994), a stimulus activated in working or short-term memory will stay there for only a few seconds; (b) is selective (Lee, Cadierno, Glass, & Van Patten, 1997; Van Patten, 1990, 1994, 1996), only one task can be attended to at a time; (c) is subject to voluntary control (Neumann, 1996), each person is free to attend to one stimulus over another; (d) controls access to consciousness (Baars, 1988; Shapiro, Arnell & Raymond, 1997), what one attends to becomes the focus of consciousness; (e) is necessary for learning (Carr & Curran, 1994; Gass, 1988; Schmidt, 1995; Van Lier, 1991; Van Patten, 1994), only stimuli attended to can become available for further processing in long-term memory.

Attention, Schmidt (2001) adds, "is not a unitary phenomenon, but refers to a variety of mechanisms … alertness, orientation, detection [among others]" (p. 3). These mechanisms will be looked at via some of the approaches in studies of attention. Most of these approaches concentrate on its relationship with awareness, as Schmidt explains that it is impossible to separate these two terms.

Schmidt's Noticing Hypothesis: The origins of attention are found in Schmidt's study of Wess, a Japanese artist whose English grammar competence became fossilised. His lack of progress was explained by a lack of attention to input. Schmidt's insights into this case, and into his own personal case study on the learning of Portuguese (Schmidt & Frota, 1986), gave rise to the Noticing Hypothesis (NH). Schmidt's (1995) NH (strong version) claims that L2 learning is possible without intention (e.g. incidental vocabulary learning) and without metalinguistic understanding, but learning is impossible without attention and its subjective experience of noticing, "noticing is the necessary and sufficient condition for converting input to intake" (Schmidt, 1990, p. 129).

Schmidt (1990) considers that the terminological ambiguity of 'consciousness' explains the constant theoretical disagreements. Thus, he calls for a distinction between three types of consciousness:

*'**Consciousness as awareness'*** within which Schmidt distinguishes different levels of awareness: 'perception' (transformation of external events into internal representations); 'noticing' or 'focal awareness' (becoming aware of an external stimulus. Different brain sections control perception and noticing, e.g. when reading, being aware of the content is noticing, while still perceiving background music or noise is perception; 'understanding' (comprehension and reflection on external stimuli, e.g. problem-solving and all forms of metacognition).

*'**Consciousness as intention'*** referring to the difference between incidental learning (without any specific intention to learn) and intentional learning (goal-oriented learning).

*'**Consciousness as knowledge'*** believing that 'knowing something' is 'being conscious of it' is a common assumption. However, Schmidt, explains that conscious and unconscious knowledge mean different things to different theorists; such differences are mirrored in their relationship to knowledge. Therefore, 'consciousness as knowledge' depends on what each theorist understands by consciousness.

Considering the above types of consciousness, noticing is the result of paying particular attention to something and entails the learner's subjective sense of awareness. This subjective experience of ´being aware of´ is what activates learning. Schmidt (1994) defines noticing as "registration of the occurrence of a stimulus event in conscious awareness and its subsequent storage in long-term memory (p. 166).

Attention without Awareness: Tomlin and Villa (1994) offer an alternative approach to the study of attention. They argue that noticing may not be as important for SLA as other attentional processes, different from awareness. They propose three attention levels: 'alertness': readiness to receive arriving stimuli; 'orientation': direction of attention to specific stimuli; 'detection': registration of selected information in memory to make it available to learning processes: hypothesis formation, testing. In this view, detection alone drives input processing. Detection is the closest level to awareness but does not entail it. Thus, they conclude that "awareness requires attention, but attention does not require awareness … awareness might come a posteriori to attention" (p. 194).

Reconciliation Proposal: Robinson (1995) states that Tomlin and Villa's view contradicts Schmidt's NH. Regarding attention levels, they suggest that 'detection' is the closest to noticing, and the level at which, according to Tomlin and Villa, learning must begin. Based on this assumption, 'detection' (and consequently learning) is possible without awareness, a key contradiction of Schmidt. Robinson seeks to reconcile these opposing views by presenting a third approach to attention, i.e. a new definition of noticing: "detection-plus-rehearsal in short-term memory, prior to encoding in long-term memory" (Robinson 1995, p. 296). Robinson explains that the activation of new information in short-term memory requires going beyond a specific level to become aware. Robinson's contribution to the noticing debate consists of downgrading the role of 'detection' in the encoding of information to short-term memory, i.e. less than that attributed by Tomlin and Villa (Leow, 1997). Robinson's (1995) proposal stems from his belief that studies of attention have disregarded the role of memory and its role in the 'control' and 'activation' of information. Tasks gain in importance in Robinson's proposal, since it is task-

processing demands that will determine the characteristics of rehearsal. Task demands call for two kinds of processing: data driven (stimuli are encoded and later brought together in working memory) and conceptually driven (processing that derives from an effort to integrate encoded stimuli within the context of other nearby stimuli). Data-driven processing leads to acquisition. Conceptually-driven processing leads to learning.

Skehan's (1998, 2003) 'limited attentional capacity model' advocates that, due to the limited nature of attention, when learners attend to one feature of language (e.g. accuracy), other features (e.g. fluency, complexity) will not be attended to. Robinson argues that some task demands do not conflict with linguistic demands, thus greater task complexity may lead to linguistic improvement. Kuiken and Vedder's (2008) L2 writing study shows that more complex tasks lead to improvements in accuracy.

Terminology Debate: Schmidt's thoughts have changed over time, perhaps due to criticism from Tomlin and Villa (1994), Robinson (1995) and other theorists. In his 2001 article, known as the weak version of the NH, Schmidt (a) changes his view of noticing from 'necessary' to 'useful but not necessary' for learning, (b) presents attention and noticing as having very low levels of abstraction and (c) sees Gass' (2008) *apperception* (when the mind identifies new linguistic data and relates it to existing knowledge), Tomlin and Villa's *detection* (when information is selected and registered in memory to make it available for learning) and Robinson's *detection plus rehearsal* (when detected information is rehearsed in short-term memory, prior to encoding it in long-term memory) as equivalents of noticing. None of these terms requires awareness. Schmidt (2001) clarifies that attention and noticing do not deal with the rules underlying linguistic utterances in the input.

Discussion of whether there can be learning without attention led Schmidt (2001) to conclude: (1) intentionally focused attention is an advantage, not a requirement for learning, (2) disagreement exists on whether all aspects of L2 input should receive attention to learn them. Van Patten (1994) suggests that different language aspects require different amounts of attention, (3) more attention leads to more learning, (4) which aspects of L2 input require attention is controversial. Carr and Curran (1994) argue that only stimuli attended to can be learned; Sharwood Smith (Internet post in Schmidt, 2001, p. 32) argues that some aspects of language are so ethereal that they cannot be attended to.

The close relationship between noticing, consciousness and awareness, different theorists' understanding of these terms and disagreement on the role of consciousness in L2 learning have led to an ongoing debate. The importance of attention in L2 learning is unquestioned among theorists. However, Reindeers (2005) points out that "they [theorists] also differ greatly in how they explain the storage and retrieval of information" (p. 31). It is agreed that attention is necessary for learning; disagreement emerges over whether attention involves awareness or not.

2.1.2 Interactionist framework: input, output, intake

The previous section looked at the cognitive aspect of the cognitive-interactionist framework of SLA. This section looks at the interactionist aspect. The interaction approach got its name from its work on speaking; its proposed constructs − input, output and feedback − occur during interaction. The cognitive-interactionist perspective looks at L2 learning as a transition from learners' L1 to the target language (L2). During this transition, various cognitive processes occur. Researchers' accounts of these processes have been turned into theories (Input Hypothesis, Output Hypothesis) and terminology (intake, interlanguage) that are central to this study.

First, for L2 learning to occur, learners' exposure to the target language is necessary. Krashen's (1985) Input Hypothesis presents input (linguistic data provided by the environment or other proficient L2 users that learners hear or receive, and from which they can learn) as a necessary and sufficient condition for learning. Krashen's claim refers specifically to comprehensible input, i.e. language that learners process for meaning which is a little above the learner's level of competence: *i+1.* Input needs to match learners' developmental level. Krashen's critics argue that he disregards several factors: learners' attentional capacity is limited (Cowan, 2001); certain aspects of language require explicit explanation (Ellis, 2003; Sharwood Smith, 1981, 1993); learners' motivation and anxiety also determine the amount of input they attend to (Dörnyei, 2005). Recently, Truscott and Sharwood Smith (2011) note that not all input is learned, thus they differentiate *input* (potentially processible language data made available, by chance or by design, to the language learner) from Gass and Selinker's (2008) *intake* (processed or analysed input that can be used for acquisition by integrating it into the learner's existing grammar). This thesis adopts these input and intake definitions. Exposure to input allows learners to observe and infer rules about L2 functioning. Based on these observations, learners produce their own L2 output. Swain's (1995) Output Hypothesis states that *output* (language produced by the learner) is both 'the result of' and 'the means for' L2 acquisition. Swain (1985) argues that to achieve native-speaker grammatical and sociocultural competence, the negotiation of meaning between interlocutors has to go beyond getting messages across. It must challenge learners to produce *'pushed output'*, i.e. to formulate accurate, coherent and appropriate messages, "just speaking and writing are not enough" (Swain, 1993, p. 160).

Swain (1985) identifies four ways in which output triggers L2 acquisition. First, it encourages 'meaningful practice opportunities' for one's linguistic resources, practice enhances fluency that will result in automaticity in L2 use. Second, it stimulates 'opportunities for noticing the gap'; output "raises learners' awareness of gaps in their knowledge ... it plays a consciousness raising role" (Swain, 1995, p. 130). Third, it fosters 'opportunities for hypothesis testing', i.e. learners' attempts to match their intended meanings with L2 resources or expressions they have. Hypothesis testing generates *feedback,* a way to draw the learner's attention to form.[3] Feedback provides input (in the form of information) that makes learners aware of the correctness (what is acceptable in L2) or incorrectness (what is not acceptable in L2) of their speech, and leads them to modify their output. In those attempts, learners self-evaluate themselves (*internal feedback*) or get teachers' or interlocutors' feedback on their hypotheses (*external explicit feedback*). Finally, output helps 'metalinguistic reflection'. Output forces learners' attention onto form and onto conscious reflection on their utterances, making them move from semantic to syntactic processing, i.e. L2 comprehension depends more on meaning and content words than on syntax, whereas in L2 production syntax and grammar accuracy become more important.

Constant cycles of learners' L2 input, output, feedback, hypothesis testing and retesting develop learners' *interlanguage*, learners' system with its own grammar and lexis that is independent of learners' L1 and L2 but with features of both (Selinker, 1972). Interlanguage constitutes an important means to explore what happens in learners' minds.

[3] Attention to linguistic features resulting from a communicative demand.

2.2 Language learning potential of writing

For a long time, SLA researchers have been interested in speaking rather than writing. Both are productive skills; however, "writing has generally been seen as having a minor role in promoting L2 development … as the result of acquisition, rather than as a facilitating factor" (Williams, 2012 p. 321). Williams acknowledges the characteristics of writing − its slow pace, its permanence and the precision in language that it requires – which are crucial for L2 development. She explains that the precision in the target language that writing requires derives from the permanence of written text, which encourages learners to write as accurately as possible. The slow pace of writing offers learners time to refer to their explicit knowledge and satisfy the precision requirement. The permanence and slow pace of writing also engage learners in attentional processes in the three stages of *L2 development* proposed by Housen and Pierrard (2005): internalization, modification and consolidation of knowledge.[4]

Writing research is divided into two areas: L2 writing and SLA. Researchers have worked independently in these two fields for many years. Ortega (2012) describes them as follows: 'L2 writing' explains the development of multiple language, written and literacy capacities. On the other hand, 'SLA' explains additional language (L2) development. SLA researchers have been indifferent to L2 writing since they consider it focuses on writing literacy, a culture-dependent variable. Conversely, L2 researchers have been indifferent to SLA since they consider it irrelevant to the development of language written literacies. The differences between these two fields are important. However, as Ortega observes, the two fields would benefit from working together. Only in 2012 did the intersection between these two

[4] Internalization (creation of form-meaning associations resulting from noticing specific input), modification (changes to initial form-meaning associations resulting from further input and feedback) and consolidation (reinforcement of form-meaning associations achieved by constant retrieval and deeper processing).

fields of writing begin to attract some interest. Bitchener, Hanaoka, Izumi, Kormos, Polio, Ortega, Storch, Wigglesworth and Williams, (in a special issue of the JSLW, 2012) called for more exchange between these two areas of writing in a special issue of the 'Journal of Second Language Writing'. This emerging change in the approach to writing is based on the *'language learning potential of writing'* (LLPW), a term attributed to Manchón but originally mentioned by Cumming (1990) as the way "composition writing elicits learners' attention to form-meaning relations that make them refine their linguistic expression − and hence their control over their linguistic knowledge" (1990, p. 483). Manchón (2011b) states that the foundations of this construct are found in Cumming's (1990) view of composition writing as a "psycholinguistic output condition wherein learners analyse and consolidate second language knowledge that they have previously (but not yet fully) acquired" (p. 483).

In 2011 Manchón conducted a critical overview of existing research (studies published between 1990 and 2010), examining different aspects of L2 writing and SLA. Her aim was to search for authors who explicitly or implicitly stated their interest in the language learning potential of writing. Manchón's (2011b) work allowed her to put forward the LLPW as a new research domain, the aim being "to investigate the writing-to-learn language dimension of L2 writing development and instruction" (p. 62). In her analysis, Manchón (2011b, p. 62) distinguishes descriptive from interventionist studies. The former examine the circumstances in which the creation of written output triggers cognitive and sociocultural processes that result in language development. The latter examine the circumstances in which experimental interventions yield short-term L2 learning. Cognitive interactionist and sociocultural theories support descriptive and interventionist studies. Section 2.1 above explains in detail the cognitive interactionist framework supporting this study. Studies supported

by the sociocultural framework contrariwise perceive L2 learning as a consequence of social interaction and mutual aid.

2.2.1 'Learning-to-write' and 'writing-to-learn'

Manchón (2011b, 2013) differentiates between two areas of investigation: the ***learning-to-write*** dimension (referring to how people learn to express themselves in writing) focuses on L2 writing research and the ***writing-to-learn*** dimension (referring to how people's engagement with writing contributes to L2 learning) focuses on SLA studies. Within the learning-to-write dimension, writing constitutes an end in itself and teaching is associated with multi-(literacy) education, i.e. literacy in more than one language (Manchón, 2013). Within the writing-to-learn dimension, writing is considered a 'means' for language learning and teaching is limited to foreign language instruction.

Ortega (2011) characterises further the above areas of writing suggested by Manchón. She explains that the learning-to-write (LW) dimension predominates in composition programmes and writing centres for college students in university English departments, mostly in the USA. Both composition programmes and writing centres help college students to become proficient writers. LW is interested in good writing and writers' development. Instruction promotes sound writing in more than one language. Research enquires into how composing skills develop in multilingual writers. L2 writing is understood as mastering linguistic-rhetorical resources to transmit the writer's authorial voice while engaged in constant L2 writing processes. This dimension is backed up by cognitive-rhetorical teaching and research philosophies from L1 composition studies. At present, Ortega (2011) explains, the original cognitive-rhetorical concerns, have moved on to "discourse-, genre-, and

corpus-oriented interests that pursue the specification of the functional-textual-rhetorical resources needed in competent writing" (p. 238).

Concerning the writing-to-learn (WL) dimension, Ortega (2011) divides it into writing-to-learn content (WLC) and writing-to-learn language (WLL). The WLC dimension is found in universities where English is the medium of instruction. There, English writing support is normally offered to international MA and PhD students. LWC focuses on the reader since they determine what content is learned and whether it has been learned. Instruction aims to develop study skills and achieve academic goals via writing activities in more than one language. Research enquires into the ways in which novice international students in higher education institutions negotiate and/or meet the demands of writing in a foreign language. The conventional conception of L2 writing (writing an academic essay, a research paper or a creative short story) is replaced by a new conception of L2 writing i.e. developing the writing ability of students who are knowledgeable in some content and who use writing to increase their expertise by learning and generating new content. Ortega explains that WLC is a result of the expansion of English for Specific Purposes and English for Academic Purposes in L2 writing.

Finally, in relation to the WLL dimension proposed by Manchón and already contextualized in FL classrooms, Ortega (2011) explains that instruction in this dimension looks at L2 learning as a way of language development rather than a way to practise grammar and vocabulary. Since it is supported by SLA principles, research focuses on learners' cognitive processes and on text. WLL introduces new constructs such as feedback for accuracy, feedback for acquisition, pushed output, languaging, noticing and processing. These days, Ortega (2011) clarifies its original cognitive and

textual concerns by moving towards social, contextual and affective factors influencing the reader-writer relationship.

Among the three areas above, this study is framed within the writing-to-learn language (WLL) dimension and its concern with feedback for acquisition (how writing − text production and feedback processing − fosters L2 development). Whereas ***feedback for accuracy*** refers to the revisions learners make to previously corrected writing, ***feedback for acquisition*** refers to learners' capacity to exploit the knowledge gained from feedback on previously corrected writing in new writing. Manchón (2011b) adds that feedback for acquisition aims at "promoting *learning* understood either as the consolidation or the expansion of linguistic resources" (pp. 57–58). Feedback for acquisition, Manchón says, demands pedagogical mediation to engage learners in deep processing promoted by explicit learning conditions. Such conditions imply different awareness levels, e.g. understanding and meta-reflection on one's noticing.

2.3 Language learning potential of WCF

2.3.1 Characterizing the language learning potential of WCF

In this section, I look briefly at other SLA theories that have something to say 'for' or 'against' the language learning potential of written corrective feedback. Bitchener (2012) suggests that SLA "theories, about why one might expect a particular independent variable (corrective feedback) to influence a particular dependent variable (accuracy), are the best place to start" (p. 349), in both research on WCF and analyses of WCF studies. Different frameworks exist in the SLA field to approach L2 development: cognitive-interactionist, skills acquisition, sociocultural, generative and processability theories are among them. Cognitive-interactionist theory

was described in Section 2.1, as this constitutes the theoretical basis of my study. Skills acquisition and the sociocultural theories are approached in this section as they acknowledge the language learning potential of written corrective feedback (WCF).

Cognitive-interactionist theory is the framework that best recognizes the learning potential of WCF. Although it does not tackle written error correction directly, its premises do. Cognitive-interactionists explain that learners receive positive and negative input from oral and written feedback. Different from speaking, feedback in a written context is always explicit (including indirect feedback); it is less likely to go unnoticed (because of the slow pace and endurance of writing) and it relies less on individual cognitive factors such as memory and attention (because of the slow pace). Therefore, as Bitchener (2012) concludes, "the language learning potential of written CF is greater than that of oral CF" (p. 351).

Although Krashen's (1985) Monitor Model belongs to the cognitive-interactionist framework, its principles do not recognize the value of feedback. First, Krashen's Input Hypothesis claims that exposure to comprehensible input is the only condition for L2 development. Second, the distinction between 'acquisition' (implicit knowledge) and 'learning' (explicit knowledge) established in his Acquisition-learning Hypothesis identifies no benefits from corrective feedback for developing learners' acquired knowledge. Finally, since his Natural Order Hypothesis suggests that learners acquire L2 linguistic structures in a predetermined order, corrective feedback and grammar teaching have no impact on L2 development. Despite not distinguishing any benefits from feedback for L2 acquisition, Krashen "does concede that teaching and CF can play an editing role in learning" (Bitchener, 2012, p. 350), i.e. in developing explicit knowledge.

Skills acquisition theory, commonly associated with DeKeyser (2007) but originally proposed by McLaughlin (1987) and Anderson (1985), presents *'learning'* as being able to do something fast and accurately. L2 learning as a skill entails the internalization of complex behaviour. This internalization is achieved by processing information in a controlled-automatic-learning sequence. Two stages in this sequence call for different types of knowledge: explicit or declarative knowledge about the skill (in the controlled stage), and implicit or procedural knowledge (in the automatic stage). A controlled-into-automatic knowledge transformation occurs in the learning stage. McLaughlin and Anderson advocate that provided there is constant practice of [and feedback on] the intended behaviour (e.g. writing), explicit knowledge (from instruction and WCF) can become implicit knowledge. Practice raises speed and performance, diminishes errors and consolidates knowledge. Feedback provides explicit knowledge, it helps learners to identify their flaws and avoid inaccurate information being proceduralized. Concerning practice, DeKeyser (2007) remarks that not only is a large amount of meaningful practice necessary for the development of L2 writing, but research on the type and quantity of feedback that is favourable during practice is necessary. Johnson (1988, p. 90) adds that skills theory implies a change in the understanding of the learning process, from 'learn → perform' to 'learn → perform → learn'. Feedback or corrective action in this interpretation of learning is an 'information providing' rather than a 'reinforcing' stage.

Concerning the potential of WCF, Bitchener (2012) signals that sociocultural theory helps in understanding its success or failure. This theory distinguishes three levels in any type of activity: (a) the reasons that stimulate it; (b) the actions resulting from the desire to do it; and (c) the circumstances or conditions that contextualize the activity's performance. Its account of learners' engagement with feedback in these

three levels contributes to the understanding of individual (cognitive, affective, motivational) and contextual factors influencing the learning potential of WCF.

It is my purpose in this section not only to characterize the language learning potential of WCF, but also to introduce a new concern of SLA researchers: the role of explicit knowledge for L2 writing development. Williams (2012) remarks that the need for precision in writing and the available time to do it call for learners' explicit knowledge in all writing stages: planning, composing and revision. Krashen (1983), McLaughlin (1987), Anderson (1985) and Polio (2012) have also drawn attention to the role of explicit knowledge for L2 writing improvement. In so doing, they pose a new question: can explicit knowledge turn into part of the developing L2 system? Or in Williams's (2012) words, "Can the creation, retrieval, or use of explicit knowledge result in a change to the developing L2 system?" (p. 325).

The key concepts underlying the study of 'noticing in WCF' have been introduced in this theoretical framework. I have introduced the complex notion of noticing and related terms. I have summarized the nature of language learning from a cognitive-interactionist perspective and the language learning potential of writing. And I have characterised the learning potential of WCF. The writing-to-learn and feedback-for-acquisition dimensions were also explained. I now move on to how these theories have been used in research.

Chapter 3: Literature Review

Feedback, as I explained in the previous chapter, is a way of providing learners with information about the correctness, or otherwise, of their output. From a cognitive perspective, errors are expected and play an important role in the learning process. Errors have been categorized in different ways. Camps, Villalobos and Shea (2012) differentiate 'local' from 'global' errors depending on how they affect the interactor's understanding: *'local errors'* (linguistic errors: the sentence sounds strange yet, the interlocutor understands totally or partially what the speaker is trying to say); *'global errors'*[5] (communication errors: the interactor misinterprets the message or considers it incomprehensible).

Ferris (2011) distinguishes 'treatable' from 'non-treatable' errors: *'treatable errors'* are related to linguistic structures that occur in a rule-governed way, e.g. subject + verb agreement, article usage, some errors in word form, punctuation, capitalization and spelling; *'untreatable errors'* are idiosyncratic and require acquired knowledge of the target language to be resolved, e.g. most word choice errors, missing or unnecessary words.

As in the case of errors, a variety of WCF techniques exist. Techniques vary depending on their explicitness (direct or indirect), focus (focused or unfocused, also known as non-comprehensive and comprehensive) or the person delivering feedback (peer feedback, teacher feedback). In this literature review, I examine previous research in three parts. Part 1 opens with a review of Truscott's claims for WCF; as his work is the reason for numerous studies, it deserves a special section to examine his claims directly. Part 2 conducts a brief chronological review of some of the best-

[5] Three possibilities may occur in this situation: (a) a proficient speaker does not understand the speaker. He detects an error but does not know what the student is trying to say; (b) a proficient speaker understands what the student is saying despite errors; (c) a proficient speaker understands something different from what the student is trying to say and does not realize that anything is wrong.

known WCF studies, it surveys studies addressing questions that have worried WCF researchers and that I share with them. Part 3 looks at the operationalization of 'noticing' in previous SLA studies, and it reviews research on noticing in WCF. A review of studies dealing with comprehensive or unfocused WCF is also conducted.

3.1 Truscott's claims for grammar correction

Disagreements have always existed regarding the efficacy of WCF for L2 learning. However, "rarely has one article in the field of language learning and teaching inspired so many empirical studies in such a short time and forced the field to examine an entrenched practice" (Polio, 2012 p. 375). Truscott's main claim is that correction of learners' grammar errors in L2 writing should be eradicated, as it does not help to improve grammatical accuracy in new pieces of writing. Truscott (1996) provides theoretical and practical arguments. The theoretical arguments include: the nature of learning being a slow process rather than a transfer of information, as correction implies; the impossibility to provide WCF at the point when the learner is ready to acquire a particular structure; the difficulty in identifying one form of WCF to treat different language areas (syntax, morphology, lexis). Practical arguments refer to (a) his disbelief in teachers' capability to give reliable feedback and (b) students' ability and determination to use feedback effectively. His disbelief in teachers does not undervalue teachers' ability to correct but rather emphasizes the complexity of error correction, e.g. a failure to perceive every single error is natural. Proofreading shows this is real, even among native speakers. Teachers may also fail to provide correct forms because, despite knowing the errors, sometimes not even experts have a clear understanding of what causes them or of what the correct usage is. In WCF, this leads to inconsistency, which may confuse learners and be detrimental for learning as it takes time away from other more productive learning activities. More importantly,

Truscott believes WCF discourages learners from constructing complex and sophisticated content. The avoidance of under-conceptualized structures and language simplification are natural learner responses to error correction and a natural strategy for successful writing (Truscott, 2007). Research has shown (Kepner, 1991; Semke, 1984; Sheppard, 1992) that learners who receive correction shorten and simplify their writing in their aim to avoid corrected structures. In Truscott's view, research findings have overestimated learners' ability to write correctly in groups where participants receive correction. As long as avoidance and simplified writing are not acknowledged as possible reasons for lack of accuracy improvement, research results are likely to be biased. Thus, Truscott calls for further research addressing error avoidance, simplification and drops in complexity. Polio (2012) argues that although the value of WCF for L2 learning is not recognised in all SLA theories, none of them has claimed it is useless or harmful for learners.

Truscott (2007) also claims that "research has found correction to be a clear and dramatic failure" (p. 271). His claim is strong and extreme for research and teaching., His statement is, however, supported by careful meta-analyses and qualitative scrutiny of existing empirical studies in WCF (see Truscott (2007) for details). In his meta- analyses he looks at least six controlled (comparing provision vs. a lack of provision of error correction) and six uncontrolled (studies offering gains in error correction without necessarily comparing them with a lack of WCF) studies dealing with the effectiveness of grammar error correction. Analysed studies include those using authentic writing samples in their measures (not grammar exercises), studies with more than one treatment and some of long duration (a 10-week period in Sheppard (1992) or one term in Polio, Fleck & Leder (1998)). Truscott concludes that the correction of grammar errors in L2 writing has 'a small beneficial effect' for

writing accuracy and that some research indicates it could be harmful for learning. Revision studies were excluded from his meta-analysis, as he makes clear that grammar correction of L2 writing is effective for revision. Revision as rewriting is part of the writing process. Truscott underlines that his claim is against grammar correction, not the provision of feedback. He believes 'error correction' has been used too broadly to include all error types. Error correction may be effective for improving certain errors of a non-grammatical type, e.g. orthographical as in Lalande (1982). Focused research on different error types is actually required, "Research literature rarely allows such [error] distinctions, so … little material is available for meta-analysis of different error types" (Truscott's 2007, p. 258).

Regarding Schmidt's Noticing Hypothesis, Truscott (1998) sees it as the suggestion that "conscious awareness (noticing) of grammar plays an important role in the [input into intake transformation] process" (p. 103). Two things should, in my view, be observed in his interpretation of this hypothesis. First, Truscott does not differentiate between consciousness and awareness (the greatest disagreement among critics of the Noticing Hypothesis). His interpretation of noticing as "conscious awareness of…" already suggests that 'noticing' entails 'conscious awareness', and that awareness is always 'conscious'. Second, Truscott explicitly refers to awareness of 'grammar' rather than to awareness of 'input' in general.

Truscott (1998) argues that: (a) the theoretical and experimental basis of the Noticing Hypothesis is weak, even in cognitive psychology, the discipline it comes from; (b) the lack of experimental support stems from the difficulty to test the hypothesis in either cognitive psychology or SLA. Since no language theory supports the hypothesis, problems in understanding what it means in SLA contexts arise; (c) because 'consciousness' has always been an uncertain and rather abstract issue in

psychology, caution is required when dealing with this construct. Despite these arguments, Truscott supports Schmidt's suggestion that consciousness is necessary for learning. He does, however, note that it is still hard to determine what attention is, when it is being focused on a certain target, and what its link to consciousness is.

In his analysis of different views about attention (as a multiple resource distributed to tasks as required, i.e. dual-task experiments – the version that divides attention into alertness, orienting and detection), Truscott (1998) considers perceptions of attention as something divisible, as a mismatch with the unitary essence of consciousness. The most interesting view on attention, in his opinion, is the one that considers it as something involving automatic (requiring no attentional resources) and controlled processes (requiring attentional resources). Consciousness characterizes controlled processes. However, as its role in automatic processes is unknown, this interesting view of attention still faces the problematic attention-awareness dilemma.

Summarizing Truscott's (1998) view, research on attention, awareness and learning is too ambiguous to explain the relationship between them. Truscott requires the Noticing Hypothesis to be acceptable to then specify: the features of language that must be noticed; what it means to notice them; a means to test awareness; a distinction between noticing and global awareness of input; and some distinction between input and understanding. Truscott (1998) puts forward a reformulation of the hypothesis, i.e. he makes the claim that "the acquisition of metalinguistic knowledge is tied to (conscious) noticing; development of competence is not" (p. 124). His reformulation of the NH aims to solve problems in the original hypothesis: the need to determine the items learners must notice and whether instruction (and grammar feedback) is efficient for L2 acquisition. By separating "noticing" from "metalinguistic"

competence, the first problem no longer exists. As for the second problem, Truscott says that research indicates that instruction is not efficient for language acquisition, though it is efficient for the acquisition of metalinguistic knowledge.

To my mind, Truscott (1996) has contributed much to SLA enquiry; for example, when he argued that empirical studies showing the effectiveness of error correction were weak, researchers were obliged to revise previous studies rigorously. Today, important flaws in different studies have been identified, especially among the initial ones. His criticisms of the Noticing Hypothesis are fair and well-founded. Far from rejecting the NH, Truscott recognises Schmidt's contribution to SLA, yet he recommends caution when dealing with consciousness. Truscott's claims go much deeper than how they are usually reported. His work, in my opinion, needs to be looked at in more detail. His claims are strong and the amount of research his claims have triggered demonstrates his arguments are solid and merit closer attention.

3.2 Experimental studies in WCF

3.2.1 Chronological review

For this chronological review, I rely on Storch's (2010) renowned critical analysis of WCF studies. I complement it with Ferris's (2004), Guénette's (2007) and Bitchener's (2012) contributions. I also delineate what I consider to be a third period.

First period (1980–2003): Most early studies had two main aims: exploring whether WCF led to accuracy improvement and comparing the accuracy effects of different types of WCF (direct and indirect mainly). Some studies had a secondary aim: (a) comparing the effects of direct and indirect WCF types (Chandler, 2003; Ferris & Roberts, 2001; Lalande, 1982; Robb, Ross & Shortreed, 1986); (b) comparing the effects of WCF and content commentaries on students' writing (Fazio,

2001; Fathman & Whalley, 1990; Kepner, 1991; Semke, 1984: Sheppard, 1992). Table 1, adapted from Storch (2010) and added to, displays the foci, findings and limitations of the main studies performed in the initial period.

As for whether WCF leads to language accuracy, Table 1 shows that of the eleven studies analysed, six support WCF leading to grammatical accuracy and five oppose it. However, flaws have been found in those six former studies: three worked with revision writing tasks, i.e. they did not include new writing tasks (Ashwell, 2000; Fathman & Whalley, 1990; Ferris & Roberts, 2001). One (Lalande, 1982) found no statistical significance and another (Sheppard, 1992) found improvement in only one of its targets (use of verbs), but not in a second one (sentence boundaries). Sheppard's study also found that grammatical accuracy and linguistic complexity improved more in a group that received content feedback than in a group that only received WCF.

Table 1 – *Research on WCF. First period: 1980–2003, Adapted from Storch (2010).*

Secondary Focus	**Study**	**Improved Accuracy? & Complementary discoveries (if any)**	**Limitations**
	Ashwell (2000)	**Yes**	Revised texts
Influence of WCF and content commentaries on students' writing	Fathman & Whalley (1990)	**Yes**	Revised texts
Differential effects of different WCF types (Two types of indirect WCF: underlining vs. underling and codes)	Ferris & Roberts (2001)	**Yes** + no significant differences in accuracy between these two types	Revised texts
Differential effects of different WCF types (Direct vs. indirect)	Lalande (1982)	**Yes** + found that students who received indirect WCF made significantly greater gains than those who received direct corrections	No statistically significant improvement
Influence of WCF and content commentaries on students' writing	Sheppard (1992)	**Yes** + improvement in the use of verbs but no improvement in sentence boundaries Group with content feedback outperformed group with only WCF	
Differential effects of different WCF types (Two types of indirect WCF: underlining vs. underling and codes)	Chandler (2003)	**Yes** + Greater gains in accuracy for students with direct WCF over those with three types of indirect WCF + indirect feedback in the form of underlining led to greater accuracy in the long term than underlining plus codes.	
Influence of WCF and content commentaries on students' writing	Fazio (2001)	**No**	
Influence of WCF and content commentaries on students' writing	Kepner (1991)	**No**	
	Polio et al. (1998)	**No**	
Differential effects of different WCF types (Direct vs. indirect)	Robb et al. (1986)	**No +** no differences for different WCF types	Revised texts
Influence of WCF and content commentaries on students' writing	Semke (1984)	**No**	

Note. The main focus of studies in this table was on whether written corrective feedback (WCF) leads to improved accuracy.

Concerning the differences between direct and indirect types of WCF, results have been conflicting: Robb et al. (1986) found no differences for type of feedback; Lalande (1982) supports indirect over direct WCF; and Chandler (2003) supports direct WCF over indirect WCF. Discrepancies are also found among studies that focused only on different types of indirect WCF. Ferris and Roberts (2001) found no significant differences. Chandler found 'underlining' resulted in more accuracy in the long term than 'underlining plus codes'.

Studies in the first research period, according to Storch (2010) and Bitchner (2012), did not succeed in showing the efficacy of WCF because design and comparability problems characterized research in this period. Design problems included the absence of a control group and new writing, inadequate writing conditions (e.g. writing at home) and no error tracking measures. Comparability problems included different populations (immersion, ESL, EFL), treatments (continuous vs. only one, grammar vs. content) and accuracy measures (mean number of errors and error ratio measures, e.g. errors/words, ratio of error free T-units to the total number of T-units).

Second period (2004–2011): Researchers' interests were the same as those as researchers in the first period. However, other WCF techniques were investigated: with or without metalinguistic explanations (Bitchener, 2008; Bitchener, Young & Cameron, 2005), with or without tutorial meetings, focus vs. unfocused (discussed in detail in the following section). All of the twelve studies that Storch (2010) analysed in this second period added a control group, included a new writing task and improved their writing conditions. Storch concludes that studies in the second period overcame previous flaws.

Table 2 presents the results of twelve additional studies analysed by Storch. Although this table looks similar to Storch's original one, I have added important information to make it useful for future reference, e.g. types of WCF compared in different studies. Research findings are presented as in Storch, i.e. under researchers' two main interests in both periods: whether WCF leads to accuracy improvement and whether different types of WCF are more effective in facilitating accuracy improvement. The findings in this period were as follows: six studies found WCF improved accuracy in immediate and delayed tests: Bitchener (2008), Bitchener & Knoch (2008), Bitchener & Knoch (2009a), Bitchener & Knoch (2009b), Sheen (2007), Storch (2009). Two other studies found that WCF improved accuracy in immediate and delayed tests, though they had some limitations: Bitchener et al. (2005) focused only on three structures; in Sheen, Wright & Moldawa (2009), accuracy was found in a delayed post-test but only for focused WCF. Three studies are of special interest: (a) Truscott and Hsu (2008), because this is the only study that reported an improvement in revised texts but not in new texts; (b) Ellis, Sheen, Murakami & Takashima (2008), because it is unusual to find no accuracy in an immediate post-test but find it in delayed post-tests; (c) Hartshorn, Evans, Merrill, Sudweeks, Strong-Krause & Anderson (2010), because it was supported by a different SLA approach, and being a non-experimental study it had no control group**.** Van Beuningen, De Jong and Kuiken's (2008) study, on the other hand, is important for this thesis because the authors and I share some design features.

Most populations in these studies were ESL adult L2 learners of intermediate proficiency. Only two of the twelve studies worked in foreign language contexts: Ellis et al. (2008) and Truscott and Hsu (2008). The 'proportion of correct usage in obligatory contexts' was the prevailing measurement of accuracy: eight (marked with

an asterisk) out of the 12 studies used this method. Concerning task types, four studies used picture stories and four used narratives; essays, letters and e-mails were less common.

The improved accuracy in immediate and delayed post-tests in half of the above studies counters Truscott's claim for the ineffectiveness of WCF for language learning. However, as these results derived mostly from focused quasi-experimental studies, they cannot be generalised. Therefore, the effectiveness of different WCF types is still unresolved.

Table 2 – *Research on written corrective feedback. Second period: 2003–2011, adapted from Storch (2010).*

Study	**Does accuracy improve?**	**Does the type of WCF make a difference?**
*Bitchener (2008)	Yes: immediate & delayed tests	Yes: Direct WCF alone was more effective than direct WCF with written explanations. **Different types of Direct WCF were compared** Direct WCF alone vs. direct WCF with written explanations
*Bitchener and Knoch (2008)	Yes: immediate & delayed tests	No effect for type of direct WCF
*Bitchener & Knoch (2009a)	Yes: immediate & delayed tests	No effect for type of direct WCF
*Bitchener & Knoch (2009b)	Yes: immediate & delayed tests	No effect for type of direct WCF
*Bitchener et al. (2005)	Yes: immediate & delayed tests But only on 2 of the 3 focused on structures	Yes: Direct WCF with individual conferences was more effective (but only for past tense & articles)
*Ellis et al. (2008)	No: immediate post-test Yes: delayed post-test	No difference (focused vs. unfocused) **Focused and unfocused were compared** This study focused on a single structure: articles
Hartshorn et al. (2010)	Yes: treatment group in post-test (new writing)	Yes: dynamic WCF (sustained, frequent) was more effective than traditional error correction
*Sheen (2007)	Yes: immediate & delayed tests	Yes: Direct WCF and written metalinguistic explanations were more effective than direct WCF alone **Different types of direct WCF were compared** Direct WCF and written metalinguistic explanations vs. direct WCF alone

Study	Does accuracy improve?	Does the type of WCF make a difference?
*Sheen et al. (2009)	Yes: immediate test Yes: delayed post-test but only for focused WCF	No differences (focused vs. unfocused) in immediate test In delayed test only, focused WCF led to improvement **Focused and unfocused were compared** This study, however. focused on a single structure: use of referential definite and indefinite articles.
Storch (2009)	Yes: immediate & delayed post-tests	Mixed findings depending on task type/length (direct vs. indirect) **Direct and indirect were compared** Direct WCF was more effective for short writing tasks (150–200 words) Indirect WCF was more effective for longer writing tasks (250–300 words)
Truscott and Hsu (2008)	Yes: revised text No: new texts The only study that reported improvement in revised texts but not in new texts	Not investigated, but it was a focused study
Van Beuningen, De Jong & Kuiken (2008)	Yes: revised texts Yes: in delayed post-tests but only for direct feedback	Yes, accuracy improvement was found in revised texts after direct and indirect feedback, but in new texts only direct feedback led to accuracy improvement. **Direct and indirect were compared**

Note. Except for Hartshorn et al. (2010), all studies included a control group and a new writing task. *WCF* = written corrective feedback. Asterisked studies used the same accuracy measure: proportion of correct usage in obligatory contexts.

Third period 2012–2015: Three events could, in my opinion, signal a third period of WCF research: (a) the call from a group of SLA researchers to look at the intersection between SLA and L2 writing, (b) recognition of the language learning potential of writing and (c) the impact of technology on writing (chats, computer-mediated feedback), teaching (writing labs, plenty of web pedagogical resources, learning management systems) and research (incorporation of latest technology in the study of language processing, e.g. eye-tracker). The SLA-L2 writing intersection and recognition of the language learning potential of writing are covered in the theoretical framework of this thesis. Technology's impact on writing, the teaching of writing and writing research is beyond the scope of this thesis. In this period, most research continues to approach focused WCF. Text length and type of writing tasks are still too controlled; the efficiency of WCF becomes even more confusing, given the incorporation of new types of WCF techniques and the wider variety of teaching contexts. New questions and research interests have emerged and previous questions have been challenged or reformulated.

3.2.2 Questions worrying WCF researchers

Numerous questions have worried and motivated WCF researchers to conduct their studies. I look at three questions that are relevant to my research.

What error types should be corrected? Pienemann (1998) and Krashen (1983) claim that only errors that learners are developmentally ready to acquire should be targeted. Truscott (1996) maintains that syntactic errors are the hardest to correct, as they are part of a complex system where most constituents are related to each other. Morphological errors in Truscott's view are less problematic but also difficult to

correct as they require comprehension of form and meaning. Ferris (2011) suggests that only treatable (rule-governed) errors should be corrected.

Concerning error categories, most studies have focused on grammatical structures, usually one, at most three. The English article system has been the most frequently analysed structure (e.g. Sheen, 2007; Bitchener, 2008; Bitchener and Knoch, 2008; Bitchener and Knoch, 2009a; Bitchener and Knoch, 2009b). Bitchener (2012) explains that some flaws in his 2005 study (he did not look at different functional uses of definite and indefinite articles) led to subsequent complementary studies: Bitchener (2008) and Bitchener and Knoch (2008, 2009a, 2009b). Simple past, articles and prepositions were also studied by Bitchener et al. (2005). Benefits were found for simple past and articles in revision and new writing. But no benefits were found for prepositions. Seldom have error categories (morphological, lexical, syntactical, orthography) been contrasted.

I found three studies comparing error types. Lalande (1982) compared grammatical and orthographical errors. She found error numbers decreased only for orthographical errors. No effects were found for accuracy improvement in lexis and prepositions. Ferris (2006) compared sixteen error types (verbs, nouns, articles, lexical and sentence errors, among others). She found a decrease only in verb errors. No effect for accuracy improvement in lexis was found. Van Beuningen (2011) compared grammatical and non-grammatical errors. She found positive effects for both grammatical and non-grammatical errors.

Bitchener's studies mentioned above targeted different grammatical structures and found significant evidence for WCF in immediate and delayed post-tests. Nevertheless, these studies cannot be accepted as evidence for the efficacy of WCF,[6]

[6] Reasons include: the rule-based nature of targeted features may have facilitated learners' processing; since the learners had a good understanding of grammar rules, WCF only triggered their consciousness;

mainly because they focused on only a few structures. The results from these studies are also in conflict with those of previous and subsequent studies focusing on other linguistic features and finding no effects for WCF on lexis or prepositions, e.g. Ferris (2006), Frantzen (1995) and Lalande (1982). Williams (2012) also emphasizes that feedback in these studies was so intensive that learners became aware of the structure being treated. To sum up, whether WCF benefits only specific language features is still unknown. In fact, Bitchener (2012) states, "the extent to which written CF can effectively target different types of error is in the very early stages of investigation" (p. 356). Further research should consider investigating not only the interaction between error categories and type of WCF, but also the relationship with other variables, e.g. the proficiency level of learners.

How should errors be corrected? Direct or indirect WCF? Direct (the error is signalled and its correction presented) and indirect (the error is signalled using codes and the learner corrects it him/herself) ways to deliver feedback have also been investigated. Pros and cons are found for both.

Direct WCF*:* Advantages include: it is immediate and less confusing than indirect WCF; it minimizes any misunderstandings learners' may have when interpreting their feedback; it allows teachers to provide rich and clear information about complicated syntactic or idiomatic errors; it offers explicit advice on learners' hypothesis testing. Disadvantages include: it might not be challenging enough for advanced learners; it might only be suitable for beginner levels; benefits depend on to what extent the linguistic target is acquired.

Indirect WCF: Advantages include: it is more challenging and engaging for learners. The problem-solving reflection it leads to may be more successful for

---

the learners' focus was more on content than on accuracy in the first writing task. See Bitchener (2012 p. 356) for details.

acquisition (Bitchener & Knoch, 2008). Disadvantages include: it is unsuitable for complicated linguistic features; it is discouraging for learners with little metalinguistic knowledge; it is not immediate.

The results from research contrasting these types of WCF are uncertain. Four studies (Guenette 2007; Van Beuningen et al., 2008, 2012; Bitchener and Knoch, 2010) found positive short-term effects for direct and indirect WCF. Direct error correction though had more significant long-term effects. Robb et al. (1986) found no differences between them. Bitchener (2012) points out that although the three studies above should be enough to support the greater benefits of direct or indirect WCF, evidence is missing to prove that direct WCF is helpful for low proficiency students and to identify the error categories that best respond to direct WCF. From an analysis of other studies comparing direct and indirect WCF (Chandler, 2003; Ferris, 2006; Lalande, 1982; Robb et al., 1986; Semke, 1984), Ellis et al. (2008) concluded that disagreements over findings were due to the different ways in which researchers had operationalized direct and indirect WCF. Ellis et al. also observe that the effectiveness of direct and indirect error correction depends on learners' grammatical knowledge of the form in question, i.e. whether it is a new or partially internalized form. As teachers are unable to know learners' interlanguage, they are also unable to select the appropriate type of WCF. Therefore, the relative effectiveness of direct and indirect WCF techniques, Ellis et al. (2008) say, "may not be the [thing] that needs investigating" (p. 355).

How much of learners' writing should be corrected? Focused or unfocused WCF? With the aim of conducting experimental research on the effects of WCF on grammar accuracy in new pieces of writing, some researchers have concentrated on focused (with a specific linguistic target) vs. unfocused (with no

specific linguistic target) WCF. Unfocused WCF is also called comprehensive and deals with all types of errors (syntax, morphology, lexis, style, cohesion, punctuation etc.).

Theoretical and practical arguments exist for and against unfocused or comprehensive WCF. For SLA theories, the success of comprehensive WCF is idealistic. Krashen's (1983) Natural Order Hypothesis and Pienemann's (1998) Teachability Hypothesis claim errors are developmental, i.e. part of learners' interlanguage. Therefore, certain errors will only be overcome when learners are developmentally ready to do so. Schmidt's (2001) conclusions about "attention" being limited, selective and voluntary are important and relevant too. Based on these theories, comprehensive WCF is unlikely to lead to L2 learning or to work in teaching practice. The systematic progress of L2 acquisition makes a focus on fewer error types more viable.

On the practical side, comprehensive WCF is said to be exhausting and stressful for teachers, and overwhelming and confusing for students. Ferris (2011) points out that "only the most competent and motivated student writers might actually study such unfocused correction and draw productive generalizations from it ... in most cases, the students will simply look at it (or not) and forget it" (p. 31).

On the other hand, Evans et al. (2010) and Hartshorn, Evans, Merrill, Sudweeks, Strong-Krause and Anderson (2010) present reasons to support the use of comprehensive WCF. Evans et al. (2010) argue that the "academic and professional worlds our students enter expect a high level of accuracy and precision … Thus, our ethical obligations as teachers should [be] to help our students write more accurately" (p. 447-448). They maintain that students are judged on their overall linguistic accuracy, rather than on their performance on a few specific structures. These

researchers make an even stronger point: no experimental evidence has been reported on the inefficiency of comprehensive WCF. More recently, Van Beuningen (2011) delineated the constraints of focused WCF for research: consideration of a limited number of errors; learners' tendency to monitor their use of the target structure; writing tasks resembling grammar exercises; and being an unauthentic correction technique.

For and against arguments are important however, though unfocused WCF studies are in fact uncommon. To the best of my knowledge, only five studies (Table 3) have investigated unfocused or comprehensive WCF. Actually, most of the studies that show evidence to support the efficacy of WCF have focused on a limited number of linguistic structures. Storch (2010) states that evidence for a limited sample of structures does not allow for generalizations about the benefits of WCF. Therefore, more research is required.

3.2.3 Five studies on comprehensive or unfocused WCF

Five studies claim to have studied comprehensive or unfocused WCF. In my analysis of these, I found common features and differences among them. All of them included a control group and a new writing task. They measured accuracy via pre-, post- and delayed post-new writing tasks. Differences included different instruments to support the studies, e.g. Ellis et al. (2008) incorporated a questionnaire to find out if learners had become aware of the tested structure. Van Beuningen et al. (2012) used a lexis pre-test to guarantee participants' language proficiency was comparable.

Table 3 – *Studies on unfocused or comprehensive written corrective feedback (WCF).*

Objective(s)	Sample & task	Treatment & analysis
Ellis et al. (2008)	**Sample**	**Treatment**
They compared the effects of **focused** and **unfocused** WCF on the accuracy of students' use of English indefinite and definite articles to denote first and anaphoric references in written narratives.	49 Japanese EFL intermediate students **Task** Narratives on picture-based animal stories	Group 1: Direct focused-WCF only on article errors (N=18) Group 2: Direct unfocused-WCF on article errors alongside corrections of other errors (N= 18) Group 3: Control (N= 3) **Analysis** Narrative writing tests (pre-test, immediate post-test and delayed post-test) scores were calculated by means of ***obligatory occasion analysis***.
Truscott and Hsu (2008)	**Sample**	**Treatment**
They were concerned with the question of how research on revision relates to the effectiveness of correction on improving learners' writing ability.	47 EFL advanced students **Task** Narrative task	They compared the performance of learners who received unfocused WCF with those who did not receive WCF. **Analysis** Errors were marked then each piece of writing was assigned an ***error rate.***
Sheen et al. (2009)	**Sample**	**Treatment**
They investigated: -The effects of focused and unfocused WCF on a single grammatical target (articles) & on a broader range of grammatical structures. -The extent to which writing practice without any WCF can lead to gains in accuracy over time.	80 adult ESL students **Task** Narrative task on a fable	Group 1: Focused WCF (N= 22) Group 2: Unfocused WCF (N= 23) Group 3: Writing practice (N= 16) Group 4: Control (N= 19) **Analysis** A series of ANOVAs with post-hoc comparisons.
Van Beuningen et al. (2012, 2008)	**Sample**	**Treatment**
They investigated: -The effect of direct and indirect unfocused WCF on L2 learners' written accuracy. -The value of WCF as a revising tool as well as its capacity to support long-term accuracy. -Truscott's claims that: correction may have value for non-grammatical errors but not for errors in grammar; students are inclined to avoid complex constructions due to error correction; and the time spent on WCF might be better spent on writing practice.	268 Dutch secondary school ESL learners, a multilingual student population. **Task** Picture-based essays on the metamorphosis of different animals.	Two experimental treatments and two control conditions. Group 1: Direct comprehensive WCF Group 2: Indirect comprehensive WCF Group 3: Writing practice only Group 4: Self-correction only **Analyses** **Accuracy**: errors were divided into grammatical and non-grammatical. Then an error ratio analysis was conducted. **Structural complexity**: text was divided into clause types (main clause and subordinate clause). Then, a subordination index analysis was used. **Lexical diversity:** Guiraud's Index was used, i.e. a type-token ratio that corrects for text length (types/√ tokens)

Objectives, target populations, tasks, treatments and types of analysis in these studies are described in Table 3. There, it is observed that accuracy measures have been various. Ellis et al. (2008) and Sheen et al. (2009) used obligatory occasion analysis; Truscott and Hsu (2008) used error rate; Van Beuningen et al. (2012) used (a) error ratio for accuracy, (b) subordination index for structural complexity (main clause and subordinate clause) and (c) Giraud's index (a type-token ratio that corrects for text length: types/√ tokens) for lexical diversity. Except for Van Beuningen et al., who used picture-based essays, all these studies used narrative writing tasks. Target populations included EFL (Ellis et al., 2008; Truscott & Hsu, 2008) and ESL (Sheen et al., 2009; Van Beuningen et al., 2012) contexts.

These five studies aimed to compare the efficacy of different types of WCF on L2 learning. At least one type of WCF in each study was 'unfocused or comprehensive'. Ellis et al. (2008) and Sheen et al. (2009) compared the effects of focused and unfocused WCF on (a) a single grammatical target (articles) and (b) the same articles along with a broader range of grammatical structures. The range of grammatical structures was different in both studies (copular 'be', regular past tense, irregular past tense and prepositions in Sheen et al. (2008); past tense, prepositions and vocabulary in Ellis et al. (2008)). Sheen et al. (2009) also explored the extent to which writing practice without any WCF leads to gains in accuracy. Truscott and Hsu (2008) were more concerned with how revision relates to correction for improving learners' writing ability. Thus, they compared a group that received unfocused WCF with another that did not. Van Beuningen et al. (2008, 2012) is the most complete study. In fact, the 2008 study was a pilot for the 2012 main study. They investigated (a) the effect of direct and indirect unfocused WCF on L2 learners' written accuracy; (b) the value of WCF as a revising tool, as well as its capacity to support long-term

accuracy; (c) three of Truscott's claims: that correction has value for non-grammatical but not grammatical errors, that students tend to avoid complex constructions due to error correction, and the time spent on WCF might be better spent on writing practice.

Regarding the findings, detailed results from each study are presented in Table 4. Considering only a first post-test, Ellis et al. (2008), Truscott and Hsu (2008) and Sheen et al. (2009) found that WCF was effective immediately after treatment for all participant groups regardless of their treatment (focused vs. unfocused in Ellis et al. (2008) and Sheen et al. (2009); underlined error vs. no WCF in Truscott and Hsu (2008). Both experimental groups improved their accuracy in a first post-test, but the underlined-error group had already outperformed the control group in this first post-test). In Ellis et al. (2008), all groups improved in a first post-test, showing no differences between them. In Sheen et al. (2009), the focused group achieved the highest accuracy gain scores for both articles and the other four targeted grammatical structures, followed by the (a) writing-practice group, (b) unfocused group and (c) control group. In Van Beuningen et al. (2008, 2012), both direct and indirect unfocused WCF led to improved accuracy over what was gained from self-editing without corrective feedback and from sheer writing practice without corrective feedback (in the 2008 study these results did not achieve statistical significance, in the 2012 study they did). Despite their differences, all of the above studies support previous research confirming that WCF is effective for editing purposes. In the case of sheer practice, Sheen et al. (2009) found that sheer practice is of value in itself, though receiving WCF is better; Van Beuningen et al. (2012) found WCF is definitely better than sheer practice.

Table 4 – *Operationalization of focused and unfocused written corrective feedback (WCF) and findings.*

Operationalization of focused and unfocused	Findings
Ellis et al. (2008) **Focused**: WCF exclusively on article errors. **Unfocused** group received WCF on article errors alongside correction of other errors (past tense, prepositions and vocabulary)	• Both focused and unfocused WCF were equally effective from pre-tests to post-tests. There were no statistically significant differences between these types of WCF in either narrative writing tests or an error correction test. • Both groups outperformed the control group, which received no correction, in a second post-test. • The results of this study contradict Truscott's claim that WCF does not affect acquisition and only assists in redrafting. There is, however, some evidence to suggest that focused WCF may be more effective in the end. This is because the unfocused group improved more than the focused group initially, but whereas the focused group continued to improve, the unfocused group did not. However, the differences between the two experimental groups were not statistically significant.
Truscott and Hsu (2008) **Focused: N/a** **Unfocused:** WCF on spelling and all grammatical errors. Errors in word choice were not included except when they involved the choice of a function word (e.g. determiners, prepositions, transitions) and could be considered grammatical problems. • Lexis and mechanics were not included except when associated with fragments and run-on sentences and failure to use a question mark instead of a period.	• In the short term, the underlined group was significantly more successful than the control group. • In the long term (one week later), the error rates from the first narrative to the new narrative in the two (underlined WCF and no WCF) groups were identical. • Thus, successful error reduction during revision is not a predictor of learning (at least for uncoded WCF).
Sheen et al. (2009) **Focused**: WCF on a single grammatical target: the indefinite article 'a' as first mention and the definite article 'the' as second mention.	• All three experimental groups (focused, unfocused and writing-practice-only groups) gained in grammatical accuracy over time in all the post-tests. This suggests that doing writing tasks is of value in itself.

Unfocused: WCF on a broader range of grammatical structures (articles, copula 'be', regular past tense, irregular past tense and prepositions)	• In the short term, the focused group achieved the highest accuracy gain scores for both articles and the other four targeted grammatical structures followed by the writing-practice, unfocused and control groups. • In the long term, the focused group outperformed the control group, whereas the unfocused group did not. The results suggest that focused WCF is more effective than unfocused WCF. • Results suggested that unfocused WCF has little pedagogical value whereas focused WCF may contribute to grammatical accuracy in L2 writing.
Van Beuningen et al. (2012)[7] **Van Beuningen et al. (2008)** **Focused: - - -** **Unfocused:** WCF on syntactic (word order, and additions or omissions of constituents) and inflectional and morphology errors (articles and pronominals)	• Both direct and indirect unfocused WCF led to improved accuracy over. -self-editing without corrective feedback (Control-Group 1) and -sheer writing practice without corrective feedback (Control-Group 2) • This result was valid for revision and new pieces of writing (post-test and delayed post-test sessions). **Accuracy:** • Direct WCF resulted in grammatical accuracy improvement in new writing, whereas indirect WCF was more beneficial for non-grammatical accuracy. **Complexity**: • WCF did not result in simplified writing when structural complexity and lexical diversity in students' new writing were measured. • The results suggest that comprehensive WCF is effective in decreasing the number of complex grammatical errors.

[7] Error categorization in one of Van Beuningen's appendices in her thesis includes more errors than the ones mentioned above. Except by lexical errors that it is known they analysed in lexical diversity, it is unclear whether the categories below (from the appendix) were also included in the reported study.
Morphosyntax: Word order error, omission of a necessary element, addition of a non-necessary element (already mentioned above) plus determiner error, referential error, inflectional errors (this one already mentioned above too).
Pragmatics: contextual errors
Orthography: capitalization, punctuation, and spelling errors
Lexicon: Word choice error. We know authors analyzed lexical diversity.

The groups' behaviour in a delayed post-test changed in all studies. Ellis et al. (2008) found no differences between focused and unfocused WCF in a second post-test. Both groups did, however, do better than the control group (but with no statistical significance). The results also show that the unfocused group only improved in a first post-test, whereas the focused group maintained continuous improvement across the tests. The authors concluded that unfocused WCF might be more effective in the long term. Both Ellis et al. and Truscott and Hsu (2008) got the same results in a delayed post-test; and from a pre-test to a delayed post-test (new narrative), the groups' accuracy in Sheen et al. (2009) (underlined-error WCF group vs. no-WCF group) was equal. In Sheen et al., in the long term, the focused group outperformed the control group, whereas the unfocused WCF group did not. These results suggest that focused WCF is more effective than unfocused WCF. In Van Beuningen et al. (2012), both direct and indirect unfocused WCF led to improved accuracy over what was gained from self-editing without WCF and from sheer writing practice, i.e. this result was true during revision and for new pieces of writing.

The researchers in the five studies above understood and operationalised 'unfocused or comprehensive' WCF differently. In Table 4, I state the implicit or explicit operationalization of this term found in each study. Sheen et al. (2009) are the only ones who explicitly present definitions of focused and unfocused WCF. Although Ellis et al. (2008) were the first to compare focused and unfocused WCF, their study was actually a focused WCF study. The study approached English articles in a focused group and the range of targeted structures was extended in an unfocused group (past tense, prepositions, vocabulary). The authors themselves state, "It might be better to characterize the differences between the two types of CF in this study as focused versus less focused rather than focused versus unfocused" (p. 367). Except

for articles, the criteria used to select the range of errors in this study are not stated. Based on the rationale Ellis et al. give for their selection of 'articles', it may be inferred that their criteria might have been the same: ease of research. Lexical and syntax errors were included in an unfocused group. This study, however, was criticised for the short, simple writing tasks included, as they do not represent common writing tasks in ESL and EFL contexts.

Truscott and Hsu's (2008) study was the second to deal with unfocused WCF. The authors claim that their approach to unfocused WCF is clearer and broader than Ellis et al.'s (2008). Their error selection was in fact less arbitrary and responded to the need for consistency and broad coverage. Different from Ellis et al.'s error selection, Truscott and Hsu (2008) did not include lexis correction, one of the hardest language features to categorize in my experience. The authors recognised that including only one type of WCF was a limitation of their study. They state that they were interested in the effects of the revision process on L2 learning rather than comparing different error correction techniques.

Sheen et al. (2009) also attempted to operationalize focused and unfocused approaches more distinctively than Ellis et al. (2008). A focused group received corrections of errors in a specific target structure: English articles. The unfocused group received corrections directed to errors in a range of linguistic structures (articles, copula 'be', regular past tense, irregular past tense and prepositions), two more structures than in Ellis et al. As can be observed, Sheen et al. (2009) and Ellis et al. (2008) called 'unfocused' what was in fact a 'semi-focused' approach. The value of Sheen et al.'s study is that it is the only study that explicitly defines what is meant by 'unfocused' WCF.

Error selection for the unfocused approach in Sheen et al. (2009) was based on the researchers' predictions of the categories' difficulty and on the reliability they offered to identify them. Sheen et al. also addressed Truscott's (1996) claim as to whether writing practice without CF leads to gains in grammatical accuracy. Despite Sheen et al.'s (2009) efforts to overcome previous research flaws, their study too had its limitations, e.g. unsystematic error correction between groups. Their focused WCF was more systematic (article errors were always corrected) than their unfocused (some errors were corrected while others were ignored) group.

Van Beuningen et al. (2012) operationalised unfocused WCF as involving 'correction of every error in students' writing'. As in previous studies, unfocused refers to a range of errors. This study, however, included more and broader categories: article, inflection, word order, omissions, additions and pronominal errors. The authors claim their error selection accounted for syntactic errors (e.g. word order and additions or omissions of constituents) in order to test Truscott's (2007) claims about the type of errors that could never benefit from WCF. The authors also explain how article, pronominal and inflection errors account for morphology errors. Although Van Beuningen et al.'s (2012) error categorisation for unfocused WCF is wider, the limitations of their study include their broad categorizations of grammatical and non-grammatical errors that make no difference to the types of features in each grammatical or non-grammatical error type. The authors themselves state, "Our findings suggest that comprehensive WCF is effective in decreasing the number of … complex types of grammatical errors. Further research is warranted, however, to identify the exact effect of comprehensive CF on separate types of grammar problems, or even different functional uses of a single grammatical feature" (p. 35). Van Beuningen et al.'s ESL context is also different from most foreign language contexts.

The authors describe their context as a naturalistic SLA environment (learners' more proficient use of language). Other limitations the authors recognise are (a) single provision of feedback, (b) specially designed writing tasks and (c) the use of error rate to measure accuracy (this method assumes gains in L2 knowledge in task 2 are as a result of treatment, but this is not necessarily so).

Some strengths and weaknesses have been outlined for the above studies. However, different criteria for error categorization and different operationalization of 'unfocused' WCF characterize all of them. The authors talk broadly about focused and unfocused groups; however, few of them offer details about the rubrics or guidelines they used to categorise grammatical subcategories or errors in general. Truscott and Hsu (2008), for instance, claim they included all grammatical errors, including word choice when it involved the choice of a function word, so it could be considered a grammatical problem. Errors in mechanics in general were not counted (except fragments, run-on sentences and failures to use a question mark instead of a period). Spelling errors were marked. Lexical errors in general were not included. Ellis et al. (2008) mixed past tense, prepositions and vocabulary in their unfocused group. Sheen et al. (2009) state that their "unfocused group received corrections on up to 8 errors involving 5 different grammatical features whereas the focused group received corrections on between 2 and 8 errors in articles only" (p. 565). However, only five grammatical structures are made explicit: articles, copula 'be', regular past tense, irregular past tense and prepositions. Whether spelling, lexis, punctuation or pragmatic errors were considered is not stated in any of the studies. Van Beuningen et al.'s (2012) grammatical error selection included six structures of grammatical error type: word order, additions or omissions of constituents (syntactic errors) and articles, pronominals and inflection errors (morphology errors). All researchers classify error

types in different ways. Some studies exclude lexis in the unfocused group (Ellis et al., 2008; Sheen et al., 2009); others exclude lexis and mechanics but include spelling (Truscott & Hsu, 2008); some refer broadly to grammatical and ungrammatical errors without specifying how lexis and other features that might have multiple functions were dealt with (Van Beuningen et al., 2012). Despite the challenges of unfocused error correction, the authors above risked addressing this pedagogically relevant WCF technique. Hence, they deserve some credit. Unfocused studies on WCF are crucial in the WCF debate. If a way to do research on unfocused WCF exists, researchers have first to agree on the operationalization of 'unfocused' and on how grammatical and non-grammatical errors are to be categorised. The type of writing task should probably be considered in this categorization too, as different writing tasks may call for different linguistic structures. Sheen et al. (2009) point out that "It is always easy to critique what researchers investigating WCF should and should not have done. The way forward is to try to investigate systematically the variables that are pedagogically relevant such as the distinction between focused and unfocused CF" (p. 567).

Before closing this section, reference must be made to Ferris's 2006 study. She approached comprehensive error correction too. Her study is, however, not included in this section because it was different in two ways: (a) it was a longitudinal study and (b) the feedback delivered was a three-draft process-oriented approach. Similar to previous studies though, error selection in Ferris's study was also based on instructors' inferences of what they considered to be the most representative errors in that context. Sixteen error categories were selected. Ferris's results show statistically significant reductions in students' number of verb errors only; the results are

insignificant for the remaining error categories.[8] Articles and sentence structure are somewhat worse at the end of the semester.

3.3 Experimental studies on noticing

3.3.1 Experimental studies on noticing in SLA

Complexity in the measurement of attention and related concepts has translated into different operationalization procedures in oral and written language. Noticing has been operationalized as (1) verbal or written self-correction of the targeted form after noticing divergence between the participant's answer and that provided by another clue, (2) the participant's comments (mmm, interesting etc.) on the targeted linguistic forms (both in Leow, 1997), (3) a learner's report, indicating that the form was new to her/him or that s/he was aware of a mismatch between the target form and her/his non-target-like production or comprehension (Mackey, 2006).

Concerning measurement procedures, online and offline procedures have been used to study noticing. As Leow (2013) explains, offline procedures take place at the retrieval stage, after data have been processed, e.g. underlining, circling, checking linguistic targets in texts, and offline questionnaires. Online procedures, on the other hand, take place at the construction or encoding stage, i.e. while learners are processing new input, e.g. language-related episodes (LREs[9]) think-aloud protocols and retrospective interviews. Leow affirms that offline procedures characterised initial

[8] Word choice, word form, verb tense, verb form, articles, singular-plural, pronouns, run-ons (a sentence that has two main clauses without connecting words or correct punctuation), fragments, punctuation, spelling, sentence structure (missing and unnecessary words, word order), informal register choices, errors in the use of idiomatic expressions, S+V agreement and a miscellaneous category.

[9] instances "in which students talk about language problems encountered while writing and (attempt to) solve them" (Swain & Lapkin, 1995, p. 378).

empirical studies (1990–1995) whereas online procedures were more common from 1996 to 1999.

Most researchers consider online or a combination of online and offline procedures as the best options to measure cognitive processes. Most measurement procedures however, and even combinations of them (Mackey, 2006), seem to have some limitations. Leow (1997) identifies some of these. Think-aloud protocols and stimulated recall require learners to report their mental processes under communicative stress, leading to underreporting. Another disadvantage is 'reactivity' (how the simultaneous verbalisation of cognitive processes influences the cognitive process one is trying to describe): participants' task performance may be affected negatively, or they may feel obliged to perform more correctly. The use of diaries and uptake sheets (Mackey) to obtain introspective data on learners' noticing superficially connects noticing to the facts that prompted them. Leow, Robinson, Mackey, Gass and Schmidt (2011) explain that input processing occurs momentarily, whereas diaries and uptake sheets last longer. Thus, there is the possibility of forgetting an experience at the time of reporting. Regarding post-exposure questionnaires (Mackey, 2006; Robinson, 1995), their usefulness depends on the learners' capacity to separate what they notice during learner-stimuli interaction from what they notice while completing a questionnaire. The latest studies (Godfroid, Housen & Boers, 2010) have incorporated cutting-edge technology, like eye-tracking "the online registration of someone's eye movements" (p. 175), into the study of noticing. These researchers use eye-tracking as a measurement procedure to quantify learner-initiated noticing of new terms in content-focused reading and its influence on vocabulary uptake.

3.3.2 Experimental studies on noticing in WCF

Truscott (1998) states that WCF is one way to make learners notice the difference between the target language and their non-target-like written production. Research on WCF, in Truscott's view, might provide evidence for the value of noticing. Similarly, Manchón's 'writing-to-learn' dimension, explained in Chapter 2, suggests that WCF research should examine how writing (composing and processing feedback) stimulates L2 learning. This section examines studies that, similar to Truscott and Manchón, see a goal for writing that goes beyond improving writing skills. The goal of writing is seen rather as a form of output that, enriched with WCF, promotes language development (Sheen, 2010). The studies analysed in this section consider the self-initiated nature of writing in order to investigate learners' self-initiated noticing.

Two types of noticing: Two types of noticing, deriving from Schmidt's (2001) Noticing Hypothesis and Swain's (1985, 1995) Output Hypothesis, are acknowledged by SLA theorists as being necessary for L2 acquisition. Both noticing functions derive from the production of language in writing which obliges learners to: ***Notice the gap*** between learners' developing linguistic system and the L2 system: this noticing is traditionally mediated through corrective feedback as incoming input. Truscott (1998) considers that, via grammar correction, WCF helps learners to notice formal features in the L2 system. Its goal is to make learners 'notice the gap'. ***Notice the hole*****:** Nassaji (2010) explains that when an interlocutor asks learners for clarification during communication, the need to make their output more precise makes learners aware of their linguistic limitations, thus a hole is noticed. Williams (2012) explains that 'noticing the hole' accounts for the 'role of output', which is believed to be stronger in writing due to its permanence and slow pace. While learners may be

unable to communicate both orally and in writing due to lack of L2 knowledge, in writing learners have the opportunity to reflect on their explicit knowledge. Composing obliges learners to notice the L2 forms they need to convey their message. Izumi (2003) explains that L2 production entails cognitive processes, such as recalling vocabulary, deciphering grammar and articulating thoughts. Learners' problems with these processes predispose them to be alert to successive input, feedback and their own output. Some researchers (Hanaoka & Izumi, 2012; Johnson, 1988; Qi & Lapkin, 2001) stress that the moment learners' struggle to communicate their meaning becomes the perfect moment for pedagogical action, as that is the moment when learners' interlanguage is open to change. Noticing in WCF studies has been studied in these two stages. Manchón (2011b) categorizes Izumi (2002) and Izumi & Bigelow (2000) in the language production or composing stage, and Adams (2003), Lapkin, Swain & Smith (2002), Qi & Lapkin (2001), Swain & Lapkin (2002), Tocalli-Beller & Swain (2005) and Watanabe & Swain (2007) in the reception or feedback stage.

Learner-initiated noticing in WCF: 'Who should noticing come from?' is a central question in WCF, SLA and noticing research. Despite, Williams's (2001) claim that "the effectiveness of FonF [focus-on-form] is ultimately determined by learners' needs" (p. 175), most WCF techniques tested in WCF research have involved teacher-provided feedback. Qi and Lapkin (2001) argue that teacher-provided feedback includes learners' failure in: attending to what teachers intend them to attend to; paying more or less attention than is requested; understanding grammar explanations; noticing language features that are considered important; looking for rules in the input when not asked to do so (Schmidt, 1993, p. 219). Teacher-provided feedback, Adams (2003) adds, is also prescriptive and discouraging for learners as their papers are returned with many confusing corrections. The above claims support

Williams's (2001) claim for the influence of learners' needs on the effectiveness of FonF. In my view, as learning and noticing take place in the learner's mind, they cannot be influenced by teachers or researchers' purposes (Hanaoka, 2007; Park, 2011). They are, rather, the result of learners' needs and their internal syllabuses (Ellis, Basturkmen & Loewen, 2001). Consequently, learner-initiated opportunities that match learners' needs (Williams, 2001) should be encouraged and explored.

The exploration of noticing in WCF has borrowed concepts from the interactionist approach and applied them to writing (Polio, 2012). Table 5 presents an overview of some studies on noticing in WCF. Objectives, treatments and main findings are included. I focus on these studies because they are the ones that most influenced my design. Table 5 shows that noticing in WCF has been operationalized as learners' externalized 'observation of' or 'comment on' features of the input (Qi & Lapkin, 2001, Park, 2011), self-reports in the form of note-taking (Hanaoka, 2007; Hanaoka & Izumi, 2012) and the amount of corrections noticed in the comparison stage of the writing task (Santos et al., 2010). These researchers differentiate noticing from their operationalization of uptake, which they operationalized as the type and quantity of accurate revisions incorporated in the participants' revised versions of their original texts.

Methods for data collection have included language-related episodes (LREs) from think-aloud protocols, retrospective interviews, learners' pointing to a specific line in a text followed by their exclamatory utterances (Qi & Lapkin, 2001; Hanaoka, 2007) and pair discussions of the feedback received on a mutually produced text

Table 5 – *Some studies approaching noticing in written corrective feedback (WCF).*

Study and Objectives	**Treatment**	**Findings**
Qi and Lapkin (2001) explored how noticing relates to composing (Stage 1) and reformulation (Stage 2), as well as the impact such noticing has on improvement of the written product in the post-test (Stage 3) of a three-stage writing task.	Case study including language-related episodes (LRE) video and audio recorded interviews. LREs were categorised as lexis, form and discourse. Each participant produced two LREs (one from each of Stages 1 and 2). The four think-aloud protocols were transcribed and analysed.	Noticing occurred in both composing and reformulation. However, the quality of noticing was different for learners with different levels of L2 proficiency. Promoting good quality of noticing was considered more important than simply promoting noticing.
Hanaoka (2007) explored what learners noticed in a four-stage writing task: learners wrote a story in response to a picture prompt (Stage 1); learners compared their original writing to two native-speaker models (Stage 2); 1st revision – learners revised their original text based on what they noticed (Stage 3); 2nd revision – learners revised their original text again after two months (Stage 4).	LREs categorised as lexis, grammar, content and other (for features that did not fit into any category). Three data categories were created: problematic features noticed (PFNs), Stage 1; features noticed (FNs), Stage 2; features incorporated, Stages 3 and 4.	Learners noticed more lexical features in the composing and comparison and revision stages. Lexical features were also incorporated in both revisions. More proficient learners noticed more features than less proficient learners when they compared their original writing with two models. The features of the models that learners noticed included those that were related to the problems that they had noticed through output. These were incorporated at a higher rate and were retained for longer than unrelated features.
Santos et al. (2010) investigated the effects of direct error correction and reformulation on noticing and uptake, as evidenced in the written output produced by learners.	Collaborative writing (composing) and individually writing (revision), guided noticing table and interviews with learners. Accuracy of each noticed item by each pair and each participant within the pairs was analysed with T-units (one main clause plus whatever subordinate clauses happen to be attached to or embedded within it).	Direct error correction and reformulation did not influence the participants' amount of noticing. The direct error correction condition led to more accurate revisions in terms of T-units and individual errors. This second measure showed statistically significant differences between both treatments. L2 writers may benefit differently from different types of direct WCF. The direct error correction condition led to similar amounts of incorporation of different types of error revisions. In contrast, when students were given reformulated versions of their texts, they tended to incorporate more revisions concerning vocabulary. Learners found it more difficult to include reformulations related to discourse issues.

Study and Objectives	Treatment	Findings
Yang and Zhang (2010) examined the effectiveness of reformulation and model text in a three-stage (composing, comparison, revising) writing task.	Pair talk recorded during S1 and S2 was transcribed in terms of content-related episodes (CREs) and language-related episodes (LREs). LREs (lexical, form, discourse) were also measured as correct, incorrect, and avoided features), interviews with participants	Students made more effort to find the proper language to express their ideas in S1 and were able to notice most differences between their original text and the reformulated one in Stage 2. Contrasting with previous studies that examined the role of reformulations, pair discussions or model texts independently, this study integrated pair discussions, reformulations and a model text at the comparison stage, which allowed participants to notice their improper language use, to be exposed to richer language input and to notice native-like language.
Park (2011) explored learners' self-generated noticing displayed by two L1 groups under two conditions: 'L2 zero knowledge' and 'some L2 knowledge' conditions.	**Marking:** participants freely underlined, circled or scribbled as they read a text. **Stimulated recall:** Using the marked items from the input-marking task as prompts, participants were asked "What made you mark this item?" Answers were recorded for each participant and later transcribed. **Post-exposure questions:** Learners responded to two questions **Question 1:** Was there anything about the text that caught your attention? Anything that stood out as interesting or strange? If so, what? **Question 2:** Did you learn anything about the language (rules, patterns) from the reading?	Under the "zero L2 knowledge" condition, both groups exhibited similar noticing patterns prompted by perceptual input properties. In the "some L2 knowledge" condition, the Japanese group noticed more input items and processed them at a deeper level: understanding. **Question 1:** L2 orthographic features were the most frequently noticed feature (more than 50% of participants from both groups commented on the shape of the characters). Other features that both groups noticed were punctuation and verb endings. **Question 2:** Verb endings was the most frequent response. Learners had no idea what these suffixes referred to. The English participants were more sensitive to repetitions, sequences and pattern changes than their Japanese colleagues.
Hanaoka and Izumi (2012) explored: overt and covert problems learners experience as they produce written output; the extent to which they notice their solutions and incorporate them in their revisions; the different roles that a model text and reformulation play in these processes.	Note-taking of problems learners noticed (a) as they wrote a story based on a picture prompt; (b) as they compared their writing with each of the two models provided during feedback. Various noticing sheets.	Learners noticed solutions to overt and covert problems and incorporated them in their revisions. The two types of feedback texts played different roles: -Model texts offered solutions to overt and covert problems almost equally; -Reformulations offered remarkable solutions to overt problems.

(Swain & Lapkin, 2007). Yang and Zhang, 2010 included content-related episodes (CRE) too. Learner-initiated noticing, a more specific concern, has included: self-reports in the form of note-taking (Hanaoka, 2007); reformulations or error corrections identified by students and recorded in noticing tables completed in pairs (Santos et al., 2010); stimulated recall or "learners' externalized observation of or comments on features of the input" (Park, 2011, p. 156).

Studies have also contributed with their designs in different ways. Different from Qi and Lapkin's (2001) who explored reformulations, Hanaoka's (2007) enriched his design by including both a model text and a reformulated text, both written by the same native speaker. This feature is important because it increased the noticing possibilities for learners. Park (2011) contributed his focus on learner internal factors that mediate noticing. He focused on the effects of learners' L1 and on their current L2 knowledge in generating noticing. Hanaoka and Izumi (2012), on the other hand, explored noticing when learners produce output (overt problems), but also when they stop articulating their messages (covert problems). Santos et al. (2010) used guiding noticing sheets and their design made the same participants experience both treatments (direct EC and reformulation).

Feedback techniques have included models (Hanaoka, 2007), reformulation (Adams, 2003; Lapkin, Swain & Smith, 2002; Qi & Lapkin, 2001; Swain & Lapkin, 2002; Tocalli-Beller & Swain, 2005; Watanabe & Swain, 2007), the provision of feedback followed by languaging, i.e. reflection on feedback (Suzuki, 2008) and tutorials (Nassaji & Swain, 2000). Studies can also be grouped into those exploring a single type of WCF (mainly reformulation, e.g. Qi & Lapkin, 2001) and those comparing different types of feedback, e.g. reformulation vs. direct error correction

(Sachs & Polio, 2007, Santos et al., 2010), self-correction vs. WCF (Lázaro, 2009) and a mixture of reformulation and editing (Storch & Wigglesworth, 2010).

Target populations, at least in WCF dealing with noticing, have started to include foreign language contexts, e.g.: Hanaoka (2007), 37 Japanese EFL learners; Santos et al. (2010), 8 Spanish EFL high school learners; Yang & Zhang (2010), 10 Chinese EFL university students; Hanaoka and Izumi (2012), 38 Japanese EFL university learners. On the other hand, Qi and Lapkin (2001) worked with 2 Mandarin background adult ESL learners and Park (2011) with 30 native speakers of Japanese and 30 native speakers of English.

Regarding their results, all the studies in Table 5 show that noticing occurred in the composing and feedback stages and with different WCF techniques (reformulation, models, direct WCF). More proficient learners noticed significantly more features than less proficient ones (Qi & Lapkin, 2001; Hanaoka, 2007); problems that learners noticed while composing were incorporated at a higher level and were also retained for longer than unnoticed features (Hanaoka). Lexical features predominated in learners' noticing and uptake when receiving WCF with models (Hanaoka) and with reformulation (Santos et al., 2010). In the 'L2 zero knowledge' condition, learners exhibited similar noticing patterns, mainly influenced by perceptual input properties. In the 'L2 some knowledge' condition, learners processed input at a deeper level: understanding (Park, 2011). Concerning overt and covert problems, Hanaoka & Izumi (2012) found that models and reformulations played different roles in learners' noticing. Model texts were efficient in solving overt and covert problems, whereas reformulations solved mainly overt problems. Direct WCF resulted in more efficiency than reformulation in terms of producing more accurate texts (Santos et al., 2010). This supports Sachs and Polio's (2007) conclusions. They

examined three different types of WCF (direct correction, reformulation, and reformulation with think-alouds) and found that direct correction resulted in more noticing and uptake than reformulation and reformulation with think-alouds in essay revisions. Despite their results, Santos et al. (2010) think the potential of reformulation must be further explored in different learning contexts and populations. The greater effectiveness they found for direct error correction over reformulation they say "may not be generalizable across learner populations and acquisitional contexts" (p. 135). Other studies (Adams, 2003; Sachs & Polio, 2007; Storch &Wigglesworth, 2010; Swain & Lapkin, 2002; Tocalli-Beller & Swain, 2005; Watanabe & Swain, 2007) analysed by Santos et al. (2010) found that collaborative writing is effective in promoting noticing the hole and noticing the gap, processing is more probable to occur in writing than in oral collaborative writing tasks, and noticing has short-term effects on learning (incorporation of corrections after WCF).

What learners pay attention to while receiving WCF has also been investigated. Manchón (2011b, p. 70) concludes that features attended by learners depend on learner- and task- related factors. The former include individual learner differences (ILDs) such as L2 proficiency, writer's goals, beliefs and motivations; the latter include time on task, form or meaning orientation of the task, stage of the writing process, learners' allocation of time to task among various constituents of the composing process. More precise information can be inferred, though, Manchón (2011b) says that research shows that "the more open the task the more focus on lexis and less attention paid to grammar" (p. 72). Swain and Lapkin (1995) emphasise that the writing stage also influences what is noticed. In their study, learners focused their attention on vocabulary in the composing stage, grammar was the focus of attention only in the editing stage.

Finally, I observe that most WCF techniques were tested in research using teacher-provided feedback, only a few studies have explored learner-generated noticing (Hanaoka, 2007; Park, 2011; Qi & Lapkin, 2001; Santos et al., 2010). Research on noticing in WCF has mainly been concerned with input. Qi and Lapkin stress that noticing should also be studied in output. Although Sachs and Polio (2007) were criticised for working merely with essay revision, most studies on noticing in WCF have concentrated on task revision (Qi & Lapkin, 2001; Santos et al., 2010; Yang & Zhang, 2010), i.e. new writing tasks have not been included.

Chapter 4: Rationale for the Present Study

The literature review in the previous chapter has shown that interest in identifying the potential of WCF to facilitate L2 learning and the type of WCF that helps learners best in their L2 writing development have triggered past and recent WCF research. As Bitchener (2012) remarks, this concern existed even before SLA theories. Two decades of WCF research have translated into much improved research designs. Recent studies have also shown that WCF improves writing accuracy [of certain language features] (e.g. Bitchener et al., 2005; Bitchener, 2008). However, as evidence comes mainly from restricted focused studies, the findings are incompatible and this make it impossible to define under which conditions WCF is effective.

Regarding experimental studies on noticing, the literature review has also revealed that the role of attention and noticing in L2 learning is well accepted by psychologists and SLA researchers. The role of awareness in learning is, however, rather controversial. Researchers' main discrepancy lies in whether noticing and awareness are separable processes (Leow, 2013) or are two sides of the same coin (Schmidt, 2001). Controversial positions on the storage and recovery of noticed information (Robinson, 1995; Schmidt, 1990, 1995, 2001; Tomas and Villa, 1994; Truscott, 2007) were presented in the previous chapter. Leow (1997) affirms that "the terminological and theoretical confusion in current psycholinguistic theory of attention in SLA … [is] mirrored in current empirical studies" (p. 471). The operationalization and measuring of noticing have been difficult for various reasons: (a) researchers' different terminology for 'noticing' and related terms, for example: Schmidt's (1990, 1993, 1994, 1995) attention, noticing, focused noticing; Tomas and Villa's (1994) alertness, orientation, detection; Robinson's (1995, 2011) noticing plus rehearsal in short-term memory; Gass's (1988) apperception; Truscott's (1998)

conscious awareness and, also from Schmidt (2001, 2010), awareness at the level of noticing and awareness at the level of understanding; (b) researchers' disagreements concerning the noticing-awareness relationship (separable or inseparable); (c) the inaccessibility of cognitive processes (Schmidt, 2001); (d) the promptness of the subjective experience of noticing (Leow, 1997); (e) Godfroid (2010) and Leow (2013) have also remarked on the methodological problems that measuring noticing entails: ***reactivity***[10] in concurrent studies and the operationalization and measurement of an ***absence of noticing***.

Agreement on terminology and further research on the noticing-awareness relationship and on finding a way to tackle 'reactivity' and the 'absence of noticing' are certainly necessary; however, the contribution of empirical studies to the understanding of attention and noticing cannot be denied. Unequivocal answers might be unattainable because, as Schmidt (2001) suggests, attention does not entail one but several mechanisms. Despite the limitations of various measurement procedures (think-aloud protocols, uptake sheets), these have also been effective in data collection. Combinations of offline (at the retrieval stage) and online (at the construction or encoding stage) procedures may continue to be the best option to measure cognitive processes. New technologies (e.g. eye-tracking) may also help shed more light on this task.

Regarding noticing in WCF research, Santos et al. (2010) stress that there are still many unresolved issues concerning how different types of WCF influence noticing and uptake, "which type of CF elicits more attention to form, which … leads

[10] A thread of research investigating empirically the effects of simultaneous data elicitations procedures e.g. whether think-aloud affects learners' cognitive processes while engaging with the L2.

to more (durable) uptake, or which [task-related] … or learner-related variables influence learners' processing of feedback" (p. 135).

The challenges presented above and Schmidt's (2001) remark that noticing is only the first in a series of cognitive processes in the input-to-intake transformation constituted important starting points for my research. I now move on to stating the gap my study seeks to fill, its contribution (4.1) and aims (4.2). Construct definitions and the identification of variables are introduced in the rationale for the design section (4.3).

4.1 Contribution of the study

Given the abundant studies in WCF, why do more research on this topic? Despite the numerous worldwide studies on WCF, two reasons led me to do more research. First, studies on this topic are scarce in Mexico. Second, current research still has some limitations: it is mainly focused, it generally includes only one piece of writing, tasks are too controlled and generally short (200 words maximum), feedback treatments are not sustained, studies are performed in controlled environments, mainly in ESL and immersion contexts. Research designs have also disregarded learners' engagement with feedback, the role of practice and the processing of feedback. Many of these limitations are explained by the demands of experimental research. Storch (2010) remarks that "in the desire to conduct more robust research, the pendulum has swung too far towards experimental studies" (p. 29). The priority in WCF studies has been on testing the effectiveness of different types of teacher-provided feedback, and little attention has been paid to learner-initiated noticing, i.e. what learners 'notice' or 'attend to' by themselves while receiving feedback. Storch and Wigglesworth (2010) explain that noticing and the "processing of feedback [are] … less … researched and understood because it is difficult to access such learner-internal cognitive processes"

(p. 305). However, Santos et al. (2010 p. 132) indicate that because of the self-initiated character of writing problems, writing becomes the perfect setting to study self-initiated noticing and focus-on-form (FonF) processes.

In my desire to contribute to the research on WCF, I have made an effort to address some of the previously mentioned limitations. Above all, an effort was made to plan *a more ecological and classroom representative design*. First, the writing task I settled on was an open, uncontrolled, self-produced, learner-centred, syllabus-based, 300-word opinion essay. Manchón (2011b) points out that since most research has been performed with controlled pedagogic tasks "it is still an empirical question whether or not the sustained engagement with complex meaning-making composition tasks can bring about learning (p. 76). A syllabus-based writing task means that writing opinion essays constitutes a syllabus requirement for the target population and, a 300-word opinion essay is longer than tasks in previous studies. Second, despite Manchón's (2011b) call for more investigation on feedback for acquisition, no previous study on noticing in WCF (to the best of my knowledge) has included *a new writing task,* i.e. analysis going beyond revision. Third, some WCF research has been performed with collaborative writing, my design explores individual writing. "Given that many forms of writing are intrinsically an individual enterprise, research findings on collaborative writing should not be taken to represent potential learning benefits of writing per se" (Manchón, 2011b, p. 76). Fourth, this research joins the few studies attempting to approach *comprehensive or unfocused WCF* (correction at all levels), the most time-demanding and frequently used WCF technique in FL classrooms.[11]

[11] The design did not succeed in this attempt and turned to semi-comprehensive EC. The experience is reported in this thesis.

Different from previous studies, the treatment in this study *adds noticing opportunities to the ECCs tested,* i.e. the treatment consisted of 'noticing-supported error ECCs or written corrective feedback (WCF)'. This treatment aimed to explore the potential of 'noticing-supported ECCs' for learners' writing accuracy. I maintain that despite the importance of noticing and attention for learning, learners' willingness to respond to feedback is usually presupposed. I suggest that providing WCF does not automatically imply that learners will pay attention. First, noticing opportunities have to be provided and attention to feedback has to be confirmed.

Finally, the study's contribution also lies in the priority given to *learner-initiated over teacher-prompted noticing*. Previous studies have looked at teacher-prompted noticing (Santos et al., 2010). I am interested in noticing which is learner-generated (Hanaoka, 2007; Park, 2011; Williams, 2001), noticing which is not influenced by teacher intervention. The design looks at learner-initiated noticing at the composing and feedback stages and extends the range of acquisition contexts by including samples of students in EFL contexts.

4.2 Aims of the study

The study has two aims, each with its own objectives:

Aim 1: Exploring the effects (if any) of different comprehensive (eventually replaced by semi-comprehensive) ECCs (DIR, REF, SELF) on learner-initiated noticing at two stages (composing and feedback) of the writing process.

- Whether different comprehensive (eventually replaced by semi-comprehensive) ECCs (DIR, REF, SELF) lead to learner-initiated noticing in the composing and feedback stages, i.e. The occurrence of noticing reports.
- If so, the study would aim to examine the amount of noticing and the error types (grammatical or non-grammatical) learners pay attention to by themselves in each

ECC at the composing and feedback stages, i.e. The number of noticing reports and error types reported as noticed.

Aim 2: Exploring the effects (if any) of the above noticing-supported ECCs at the feedback stage on learners' writing accuracy in the rewriting and new writing stages.

- Whether the input provided by different (the ones above) noticing-supported ECCs at the feedback stage has any effects on learners' writing accuracy in the rewriting and new writing stages, i.e. The occurrence of effects of noticing-supported ECCs in the rewriting and new writing stages.
- If so, the study would aim to examine what error types (grammatical or non-grammatical) are more amenable to correction in different noticing-supported ECCs, i.e. The error types learners corrected by themselves in different ECCs.

4.3 Operationalization of constructs and identification of variables

To work towards the aforementioned aims and positioned within the cognitive-interactionist framework of SLA, and in relation to the noticing debate, the language learning potential of writing and the writing-to-learn and feedback-for-acquisition dimensions (Manchón, 2011a, 2011b), I designed a four-stage (composing/ error correction-noticing/ rewriting/ new writing) study including three different error correction conditions and a control group.

'Direct error correction' is a WCF technique where all errors types are signalled and its correction presented. It was selected because, despite its disadvantages (isolated corrections, unclear comments, lack of learner's engagement with cognitive processing and emphasis on faults), it a widely used WCF technique in FL classrooms. Evidence for its efficiency is central to pedagogy and justifies further research.

'Reformulation' is a WCF technique that corrects errors and mistakes in a text, maintaining its content but offering a native speaker's (or proficient L2 speaker's) version (Cohen, 1983; Johnson, 1988). It was included in this study as it is a learner-centred (content and context are created by the learner), tailor-made (features noticed are those closer to learners' interests and needs) feedback technique. It is one of the least intrusive WCF techniques and has proved to be effective in promoting learners' noticing (Allwright, Woodley & Allwright, 1988; Yang & Zhang, 2010). Reformulation was operationalised in this study by having a native writer of the target language rewrite the learner's text, maintaining his/her ideas, making it as native-like as possible (Cohen, 1983, p. 4). Being the first time I used this technique, I opted for having native speakers reformulate learners' essays. I wanted the reformulation experience to be as close as possible to its definition. Three native speakers rather than one were necessary because of the task length and the number of essays to be reformulated within a short period of time. Reformulators' previous training and their participation in a pilot experience (see pilot study chapter) compensated for these decisions.

'Self-correction' is an ECC that implies no external explicit feedback, i.e. the learner self-corrects his or her production after monitoring their own output. It was selected because it triggered noticing during a previous pilot study. Then, learners were able to notice their own errors immediately after they received their original text, i.e. even before feedback was provided. With "self-initiated noticing" being the focus of my research, I considered it important to include a self-correction condition.

The three experimental groups described above, plus a control one, did an opinion essay writing task. Topic, length and genre were controlled. Existing studies on noticing in WCF have only worked with composing and revision tasks (Qi &

Lapkin, 2001; Santos et al., 2010; Yang & Zhang, 2010). This study includes a new writing stage. Assessing the effects of WCF beyond revision accounted for Manchón's (2011b) ***writing-for-acquisition*** dimension, i.e. the way writing − text production and feedback processing − fosters L2 development. ***Feedback for acquisition*** refers to learners' capacity to exploit the knowledge gained from feedback on previously corrected writing in new writing. The writing-for-acquisition dimension opposes the feedback-for-accuracy dimension by referring to the revisions learners make to previously corrected writing.

The design can be explained in two parts, each corresponding to one of the two aims of the study. Table 6 was built to support its comprehension.

Table 6 – *Visual rationale of the design.*

Aim	**Effects were observed at two moments…**	**What I was looking for**	**Type of analysis**
Explore the effects (if any) of different comprehensive (eventually replaced by semi-comprehensive) ECCs (DIR, REF, SELF) on learner-initiated noticing at two stages (composing and feedback) of the writing process. **Independent variable:** Semi-comprehensive ECC **Dependent variable:** Learner-initiated noticing (occurrence, amount and error type) reported as noticed on NS-1 and NS-2	Composing (NS-1) Accounting for the value of output (composing) to encourage noticing the hole (Nassaji, 2010)	Occurrence and number of noticing reports, as well as error types reported as noticed on NS-1	Problematic features noticed and reported on NS-1 (qualitative analysis)
	Feedback (NS-2) Accounting for the value of input (feedback) to encourage noticing the gap (Nassaji, 2010)	Occurrence and number of noticing reports, as well as error types reported as noticed on NS-2	Features noticed and reported on NS-2 (qualitative analysis)
Explore the effects (if any) of the above noticing-supported ECCs at the feedback stage on learners' writing accuracy in the rewriting and new writing stages. **Independent variable:** Noticing-supported ECCs at the feedback stage. **Dependent variable:** Learners' writing accuracy (grammatical and non-grammatical) in the rewriting and new writing stages.	Rewriting (post-test) Accounting for feedback on accuracy (Manchón, 2011b)	Occurrence of effects (of noticing-supported ECCs) and error types (grammatical and no-grammatical) in essays 2 and 3.	Error rate analysis: essay 2 (quantitative analysis)
	New writing (delayed post-test) Accounting for feedback on acquisition (Manchón, 2011b)		Error rate analysis: essay 3 (quantitative analysis)

Note. ECC = Error Correction Condition; *NS* = Noticing Sheet; *DIR* = Direct group, *REF* = Reformulation group, *SELF* = Self-correction group.

4.3.1 The relationship between ECC and noticing (Aim 1)

The first aim of the design dealt with the relationship between ECCs and learner-initiated noticing. The design explores the effects of three comprehensive ECCs (direct error correction, reformulation and self-correction) on learner-initiated noticing in two stages (composing and feedback) of the writing process. In aiming to devise a design representing a classroom, I selected comprehensive error correction (eventually replaced by semi-comprehensive ECC), as it is the most frequent and time-demanding error correction technique in FL classrooms. It was my goal to find out whether the ECCs tested (independent variable) led to learner-initiated noticing in the composing and feedback writing stages. If so, I wanted to know how different (concerning the amount of noticing and error types) learner-initiated noticing was at these two stages of the writing process (dependent variable).

This first aim of my design demanded two conditions. On the one hand, it meant offering learners noticing *opportunities while composing (output) and while receiving feedback (input)*. I used noticing sheets (an online method) to achieve this. Noticing Sheet 1 (henceforth NS-1) was offered to learners while writing essay 1. Noticing Sheet 2 (henceforth NS-2) was offered to learners while looking at their feedback on essay 1. The study implemented an extended view of Schmidt's (2001) weak version of noticing, i.e. the view where he changed his conception of noticing from 'necessary' to 'useful but not necessary' for learning. Schmidt's weak version of his NH presents ***noticing*** as an attentional process by which the mind identifies new [or problematic] linguistic data in the input [and output] and relates it to existing knowledge. The extended version of this definition includes two additions. The first is that "new [or problematic] linguistic data" recognises that data identified by the learner does not necessarily have to be new, they might be familiar but partially

learned. The identification of problematic linguistic data is also the purpose of encouraging noticing in the composing stage. The second addition, "in the input [and output] …", accounts for Swain's (1985, 1993) and Qi and Lapkin's (2001) conviction that noticing also takes place in the output. Problematic and/or new linguistic data are selected and registered in short-term memory to make them available to learning processes. In Schmidt (2001), both attention and noticing deal with the surface structure of language. The reason for this characterization is, according to Schmidt and Adams (2003), to differentiate metalinguistic awareness (also called understanding) from noticing.

Godfroid et al.'s (2010) definition of learner-initiated noticing was also adapted to fit the definition of noticing implemented in this study. ***Learner-initiated noticing*** is defined as "episodes when learners pay attention, by themselves in the absence of any external intervention, to new [or problematic] linguistic data in the input [and/or output], and relates these to existing knowledge". Both noticing and learner-initiated noticing were operationalized as learners' written reports of their language difficulties (L2 production limitations and corrections) reported on noticing sheets. This operationalization allowed me to account for (a) problems as limitations during language production (output), i.e. noticing the hole, and (b) problems as corrections in feedback (input), i.e. noticing the gap. Nassaji, (2010) explains that ***noticing the hole*** is a type of noticing that occurs when an interlocutor asks learners for clarification during communication. The need to make their output more precise makes learners aware of their linguistic limitations, thus a hole is noticed. ***Noticing the gap*** refers to the distance between learners' developing linguistic system and the L2 system: this form of noticing is traditionally mediated through corrective feedback as incoming input.

The second condition that the first aim of my design demanded was *making sure learners' noticing was spontaneous,* i.e. self-motivated and not activated by teacher's intervention or task characteristics. This requirement justified:

Minimal instructions on noticing sheets: Hanaoka and Izumi (2012) used similar NSs in their research. However, their detailed instructions are suspected to have influenced what learners noticed. The instructions in my NSs, therefore, were limited to inviting learners to specify whatever they considered to be an L2 problem while writing (NS-1) or what attracted their attention while looking at their feedback (NS-2).

Allowing learners to leave the noticing sheet blank if they wanted to: This instruction in both NSs was necessary to avoid forcing learners to report. Despite being a hazardous decision (I might have no data to analyse), it was also necessary to ensure noticing (if any) was learner-motivated.

4.3.2 Relationship between noticing and accuracy (Aim 2)

The second aim of the design dealt with the relationship between noticing and accuracy. It embraced the treatment (ECCs supported by noticing opportunities during feedback), which became the independent variable in the second part of the design. The effects of noticing-supported ECCs at the feedback stage (independent variable) on learners' writing accuracy (grammatical and non-grammatical) in the rewriting and new writing stages (dependent variable in the second part or aim of the design) were explored. NS-2 provided information for the independent variable; essays 1 (composing), 2 (rewriting) and 3 (new writing) provided information for the dependent variable.

In most WCF studies, noticing has been assumed to occur spontaneously after feedback delivery. My hypothesis is that input provided by different types of WCF

must be accompanied by noticing opportunities that ensure the learner will direct their attention to the input provided.

The difference between *feedback* and *error correction condition* in this study is important. I define ***feedback*** (the what) as input (in the form of information) about the correctness (what is acceptable in the L2) or incorrectness (what is not acceptable in the L2) of learners´ linguistic performance, speech or output. Feedback aims to make learners correct their inaccuracies and modify their output if necessary. Feedback may be: *(a) external* (information is provided by someone else, e.g. the teacher, more advanced interlocutors, L2 native speakers or the environment), or *(b) internal* (information resulting from learners attempts to achieve correctness by themselves, i.e. information is self-provided or self-initiated as learner's self-correction). A variety of WCF techniques exist depending on how feedback is provided: its explicitness (direct or indirect), its focus (focused or unfocused, also known as non-comprehensive and comprehensive) or the person delivering it (teacher feedback, peer feedback, self-correction).

Error correction condition (the how) refers to the specific techniques (explicitness, focus, person delivering it) used to provide feedback in each experimental group. The three ECCs included in this study were direct and comprehensive but differed in terms of the person delivering the corrections). Therefore, error correction condition was operationalised as internal (learner self-provided) or external (provided by others) information about learners' linguistic performance in their written essays.

ECCs differed concerning the person delivering feedback: the researcher in DIR-ECC, native speakers in REF-ECC, and learners themselves in SELF-ECC. In the REF and DIR groups, feedback was external, i.e. provided by others. In the SELF-

correction group, feedback was internal, i.e. provided by learners' themselves, as self-provided or self-initiated feedback, which I define as information about the correctness or incorrectness of a learner's linguistic performance that derives from the learner's self-correction.

To investigate the first aim of this study (the relationship between ECC and noticing), learners in the three experimental groups were encouraged to notice things from their own written output (essay 1) in the composing stage. However, in the feedback stage, learners in the DIR and the REF groups were encouraged to notice from the ***external explicit feedback*** provided by others (the researcher in the DIR group, reformulators in REF-G). Learners in the SELF group were encouraged to notice directly from their own written essays without any external feedback (***self-provided or self-initiated feedback***). To investigate the second aim of the study (the relationship between noticing and accuracy), the effects of noticing-supported ECC were measured in the learners' accuracy performance across the three essays. ***Accuracy*** was operationalized as the percentage of correct usage of grammatical and non-grammatical features.

Concerning data analyses (only introduced here and further explained in the data analysis chapter): (a) the effects of ECC on learner-initiated noticing in the composing (NS-1) and feedback (NS-2) stages were analysed qualitatively using Nassaji's (2010) terminology: problematic features noticed and features noticed, respectively; (b) the effects of noticing-supported ECCs on learners' grammatical and non-grammatical accuracy were analysed quantitatively across the three essays.

Finally, the reader should note that I use ***rewriting*** rather than ***revision*** in the third task stage (composing/ error correction noticing/ rewriting/ new writing). Revision might be understood as learners looking at their corrected or reformulated

essay 1 to write essay 2. This was not the case and will be further explained in the procedure section. The rationale for the supplementary exit questionnaire will also be explained in the data collection methods in the main study chapter.

SECTION 2: PILOT STUDY

Chapter 5: Method

5.1 Participants and objectives

A pilot study, involving 10 students (6 males, 4 females), was done at CELE, UNAM with three objectives: training reformulators and testing the reformulation technique; testing the data collection methods (writing task, noticing sheets and questionnaire); performing error analyses of learners' written essays.

Training reformulators meant introducing the study and the technique to native speakers and preparing them for their role as reformulators. Testing the reformulation technique with CELE students was also necessary. Although reformulation is one of the best techniques to promote noticing, it is little known in Mexico and seldom used at CELE. Testing the data collection methods (writing task, noticing sheets and questionnaire) was essential to make sure the instructions in each instrument were clear, to test the time for each task and to test the efficacy of each instrument to collect the data I was looking for. Performing error analyses of learners' written essays aimed to find out the type of language generated by the writing task and identify learners' linguistic needs. No target linguistic features were predetermined as I aimed to explore comprehensive feedback and wanted learners' essays in the pilot study to define what the learners' linguistic needs in this specific task were.

5.2 Implementation and results

5.2.1 Training reformulators

Three CELE teachers (American, Australian and British) participated as reformulators. Common features among them were a bachelor's degree, a CELE teacher training diploma, ten years' teaching experience (two of them), three years as British reformulator, fluent speakers of Spanish.

Implementation: Training sessions were prepared to: introduce the reformulation technique and study to the reformulators; explain what they were expected to do; familiarize them with reformulation samples from different journal articles; trigger questions on what and how much to reformulate using samples of learners' opinion essays collected in previous teaching. In this study, reformulation is defined as a WCF technique that corrects errors and mistakes in a text, maintaining the content of the text but offering a native speaker's (or proficient L2 speaker's) version (Johnson, 1988). Overall agreement on how to reformulate was the main outcome of these training sessions. Reformulation would: concentrate on accuracy (of grammar, spelling, vocabulary); be restricted to the sentence and paragraph levels, i.e. paragraph order would not change; enhance the use of formal language and connectors; use the English variety selected by learners provided only one was used consistently throughout the essay. Reformulators and the researcher inferred that focusing on these areas without changing paragraph organization would facilitate learners' identification of errors in the feedback stage and avoid text appropriation (learners' resistance to native speakers' modifications). Reformulators were requested to respect the content in original text and push output, i.e. help learners formulate accurate, coherent, appropriate messages that go beyond getting the message across. At the end of the composing stage, each learner's essay was photocopied and distributed among the reformulators who had ten days to reformulate them. The day before the students' feedback session, the reformulators and the researcher discussed the experience.

Results: The reformulators acknowledged the task had been more challenging than they expected. Their first attempt to reformulate was to rewrite learners' sentences using correct English. However, they claimed reformulation at this level

corrected learners' errors but did not make essays native-like. Johnson's (1988, p. 92) differentiation between 'reconstruction' and 'reformulation' became relevant here. ***Reconstruction*** corrects errors and mistakes in a text to make it error-free. In doing so, it helps the learner understand his/her mistakes. This, however, does not imply that a native speaker would express the same content in the same way. ***Reformulation***, on the other hand, maintains the content of the text but offers a native speaker's (or proficient L2 speaker's) version. A clearer understanding of what reformulation would imply in this study and its differences from reconstruction constituted an important result. Reformulation was operationalised in this study as having a native writer rewrite the learner's text, maintaining his/her ideas, but making it as native-like as possible (Cohen, 1983, p. 4). However, reformulating open tasks like opinion essays implied more sophisticated, refined rewriting going beyond reconstruction.

Learners' differences in English proficiency and writing ability became evident during reformulation. Some essays were well written, thus rewriting was possible (Student-4); other essays were cumbersome to reformulate, especially those where thought processes began in Spanish and were directly translated into English (Student-10). Previous research studies (e.g. Park, 2011) suggest that learners' L2 proficiency influences the benefits of reformulation. Essays and noticing sheets in the pilot study confirmed that the better the learners' L2 proficiency was, the more noticing and the less reformulation were required. As a pilot study with no measurement of learners' L2 proficiency, this is only a speculation.

Reformulators identified spelling as learners' main recurring problem in their essays. Adding numerous connectors (to produce cohesive wholes), vocabulary choice and word repetition were frequent problems too. Reformulators said "cases when students clearly think in Spanish before using the equivalent in English deserve

attention". One reformulator commented that the absence of commas or full stops in paragraph-size sentences was remarkable.

5.2.2 Testing data collection methods

Writing Task

The pilot study sought to test the underexplored genre of the opinion essay. Picture descriptions and picture-based stories prevail in WCF studies. Reformulation is frequently used with these picture-based tasks, as directed tasks are preferred for the control of what learners write. The opinion essay was selected in response to CELE syllabus requirements. Agreement with the school curriculum meets learners' expectations and facilitates administrative and academic support from authorities and teachers. Writing opinion essays also constitutes a crucial skill for EFL university students.

Implementation: Piloting writing prompts was also decisive to determine the essay topic, its length and the time allowed for writing it. Learners assessed these task features via multiple-choice questions included on NS-1 during the pilot study.

The essay topic had to be one that encouraged writing and required no previous investigation, so that learners could write it by depending on everyday knowledge. Three writing topics were tested and the most motivating one for learners was chosen for the main study. Concerning essay length, previous WCF studies on learners with similar L2 proficiency have been short, e.g. 250-word texts (Bitchener, 2008), a paragraph of about six sentences (Hanaoka & Izumi, 2012) or 120–150 words (Yang & Zhang, 2010). Thus, I aimed to explore longer texts; a 450-word essay was tested in the pilot study.

Results: (a) *An underexplored genre*. Open-ended tasks such as the opinion essay triggered spontaneous, unrestricted use of language. Learners were free to say

what they wanted and display their full (good/ bad) writing ability. The task type and genre, however, posed new challenges in terms of the 'what' and 'how much' to reformulate. Reformulation of opinion essays opened up various ways of reformulating the same ideas, dealt with sociocultural features in essay writing and pushed the reformulators to tackle text organization. The task made the researcher and the reformulators question the word reformulation: reformulation, reconstructing, rewriting? By a native speaker or a non-native speaker, teacher and/or researcher?

(b) *Writing topic*. The writing topic was selected for its interest, provocative nature and familiarity to the students. The topic should not require previous investigation so learners could write about it based on their experience and everyday knowledge. Two topics were tested in previous teaching with eighteen students, both proved equally motivating. One was used for this pilot study and assessed again by students.

(c) *Essay length.* Essay length was found to be a task feature significantly influencing the success of this study. Word count was the most tested feature in teaching (1,450–1,500) and the pilot study (400+/-50). I concluded that the more learners write, the less time they have to pay attention to their outcome and proofread their work. After two trials, I concluded essays must be restricted to 300 words, longer than current studies but manageable for research. The word limit in the main study had to be strictly respected to allow for careful writing, proofreading, word count, quality of noticing and a reasonable amount of feedback to process. The word count requested in international examinations (TOEFL and IELT writing sections) matches the new count set for this task. Nine out of ten participants judged the text length to be adequate.

(d) *Time restrictions.* Based on previous teaching and the pilot study, a two-hour class was found adequate for the various tasks required from learners: drafting, writing a

final legible version, proofreading and word count. Attention span also emerged as a factor to consider. Two hours of concentrated writing is enough to avoid learners becoming tired. Regarding time, previous studies (Sachs & Polio, 2007; Santos et al., 2010) have shown that setting a time limit for the composing and comparing (with its reformulated text) stages hinders learners' performance in subsequent writing tasks. Thus, the pilot study set no time restriction, other than the two-hour class period for all task stages.

Noticing sheets

Noticing Sheets (NSs) are records of participants' perceived problems (composing stage) and errors (feedback stage) while writing. These records were used here as evidence of learners' noticing. Noticing sheets were selected because (1) they are a way to capture learners' attentional processes in real time, in a non-disturbing way and reasonably easily compared with other methods; (2) they also allow learners to analyse input autonomously.

Implementation: Significant changes were made to the noticing sheets. The pilot study originally aimed to test three NSs (one during the composing stage and two during the feedback stage). Two different noticing conditions (teacher-prompted and learner-initiated noticing) were to be tested in the treatment or feedback stage. However, the long length of learners' essays (that translated into a large amount of feedback to analyse and the fact that sentences in reformulated essays were not necessarily in the same sequence as in learners' original texts) obliged me to scrap the teacher-prompted noticing sheet. This cancellation and the positive results (encouragement of learners' feedback processing) obtained with the learner-initiated noticing sheet led me to include a single noticing sheet during the treatment (referred

to as NS-2 in the main study), plus the already considered noticing sheet for the composing stage (referred to as NS-1 in the main study).

(a) Noticing sheet-1 was used in the composing stage to encourage output noticing. Learners jotted down their L2 problems as they wrote their essay. Hanaoka and Izumi (2012) used similar sheets. However, their detailed instructions are suspected to have influenced what learners noticed. The instructions in my noticing sheet-1 were limited to inviting learners to specify whatever they considered to be an L2 problem while writing. Leaving the sheet blank and using Spanish were allowed.

(b) Noticing sheet-2 was designed for the comparison stage, when reformulated essays were compared with learners' original writing. This sheet encouraged input noticing and processing of feedback. Instructions asked learners to select at least five language features they considered important, and I added the question: "Is there anything in your reformulated text that attracts your attention? Anything that stands out as interesting or strange? If so, what?" adapted from (Park, 2011).

Results: Noticing sheet-1 was successful, instructions asking learners to report specific problems rather than general language areas worked well and facilitated data analysis when compared with NS-2. NS-1 encouraged learner-initiated noticing, and according to the questionnaire results (Q6) and NS1-NS2 comparison, it also predisposed learners to attend to their feedback. Individual differences emerged from the data analysis of noticing sheet-1. Students 5, 6, 8 and 9 were reflective and seemed self-aware of their writing weaknesses. Student-2, conversely, wrote vague comments and showed little reflection.

Noticing sheet-2 was successful too. It triggered learner-initiated noticing and learners' own selection of the features they wanted to improve. A plus point of

noticing sheet-2 was the metacognitive reflection it led to. Learners did not only categorise errors (connectors, discourse markers), but also reported actions to overcome them, e.g. using reformulation more frequently, copying texts to improve spelling etc. Noticing sheet-2 generated plenty of valuable data. All problematic features noticed in noticing sheet-1 corresponded with features noticed in noticing sheet-2. This was also confirmed by Q6 in the questionnaire (all students confirmed that they were predisposed to paying attention to their feedback). Learners' noticing in the comparison stage was more affluent than in the composing stage.

Learners' reports on noticing sheet-2 also showed differences in feedback processing. Student-3, for instance, categorized his errors as adverbs, adjectives, idioms; Student-6 identified his weakness as lack of connectors and cohesive devices; other students simply reported errors with no further analysis.

Questionnaire

Implementation: A questionnaire was designed to investigate learners' response to reformulation and the importance or unimportance they attached to noticing. Qs 1–4 in the questionnaire investigated learners' response to reformulation; Qs 5–6 identified factors influencing noticing; Q7 explored the effectiveness of noticing for reformulation and error correction. Spanish was allowed to avoid L2 restrictions. The questionnaire elicited quantitative (multiple-choice) and qualitative (open-ended questions) data.

Results: *Learners' response to reformulation.* Eighty per cent of the students responded positively, 20 per cent negatively. The former thought reformulation was an excellent, instructive technique. Learners enjoyed tracing and making sense of their mistakes by themselves, appreciated not making errors the centre of attention, and thought vocabulary and connectors improved their essays. Student-5 said, "If

reformulation was used in the classroom it would be much more effective." Twenty per cent, however, found reformulation impersonal, "Except for spelling mistakes, I do not understand several of the changes made" (Student-3). Student-10 explained s/he felt someone had done her work for her and suggested their corrected errors be highlighted.

Advantages of reformulation. Advantages mentioned included: prompting reflection, encouraging improvement (working on their own content was engaging), an optimal way to increase vocabulary, giving exposure to their own well-written essays. The main disadvantage was learners' limitations in understanding particular errors or changes, "I need someone who tells me how I can improve what I wrote" (Student-5). All the students said having native speakers reformulate their essays motivated them to pay attention to feedback, because: (a) they wanted to know whether native speakers understood what they wrote; (b) they were curious to see how native speakers expressed what they tried to communicate; and (c) native speakers' corrections were more reliable.

Finally, 90 per cent of the students enthusiastically accepted the reformulations because their essays became clearer and more appealing and interesting to read. They appreciated phrases adding emphasis to their ideas and said reformulation was a new way of learning. Ten per cent of students were indifferent, saying it was simply another feedback technique.

Factors influencing noticing. Students' criteria to select the errors they wanted to improve included: the most repetitive sentences or expressions that were completely different from their own sentences, changes they liked the most (those that made their essay clearer and more coherent) and errors that could hinder getting their

message across. All students claimed that thinking about their problems in the composing stage had predisposed them to paying attention to feedback.

Effectiveness of reformulation for noticing. Seventy per cent of students considered reformulation was more effective than traditional direct correction, because reformulation showed them their strengths, corrected their errors and added options to express their ideas; 20 per cent (those who found reformulation impersonal) thought error correction was more effective, being more concrete and allowing them to concentrate on specific problems; 10 per cent preferred a different technique.

Essays' language analyses

Implementation: An error frequency analysis of students' essays was performed aiming to identify (a) the kind of language opinion essays led to and (b) learners' main linguistic problems in this specific task. The analysis began by correcting essays and categorizing errors. Categories were created and modified from data arising. When categories were set, errors were revised and 1 point was added to the category each error belonged to. The error frequency was determined per category (no. of errors in each essay, no. of errors in total, relative proportion for each type of error). Data were coded by the researcher several times (intra-rater reliability).

Results: Error frequency analyses of learners' essays were performed four times. First, I categorized errors in lexis, discourse and form (spelling included). Form came top (73.4% of 304 total no. of errors), followed by lexis (19.8% of 82 total no. of errors) and discourse (6.8% of 28 total no. of errors). Then, I worked with form only where spelling was the most frequent problem. Finally, I considered grammatical features only. Top linguistic features and relative proportions were: syntax (11.3%), prepositions (7.9%), subject-verb agreement (7.5%), verb + preposition/ no preposition (6.7%), omission of subject (6.3%), third person singular (6.3%). An

analysis of grammatical features (spelling, lexis and discourse devices excluded) suggests the greatest difficulty occurred with syntax. However, as (a) the syntax category was too broad and (b) these results might have been affected by different essay lengths, a second more careful analysis was performed. The revised analysis considered the first 350 words in each essay and focused on grammar components (morphology and syntax). A categorization of grammatical structures emerging in learners' essays with concrete examples and types of grammar components they belong to was done (see example below, Table 7). The complete final categorization is presented in the main study chapter.

Table 7 – *Example of grammatical error categorization in learners' essays.*

Structure	**Example of learner error**	**Grammatical type**
Irregular plurals	Instruments in our lifes	Morphological
Subject +verb agreement	People is used to seeing...	Syntactical

The results of the revised analysis were: syntax 52.4% of a total of 108 errors, morphology 24.4% of a total of 50 errors, lexis 22.9% of a total of 47 errors. The top five errors were in syntax: unnecessary definite article (UDA), omission of subject, S+V agreement, prepositions, verb + preposition/ no preposition. Prepositions, subject +verb agreement, verb + preposition/ no preposition and omission of subject deserve special attention as they were syntactical features emerging in the original and revised analyses. These results offered quantitative data about structures merely inferred by the CELE teachers.

Chapter 6: Implications of Results for the Main Study

The results from the pilot study were encouraging: open tasks proved optimal for learner-initiated noticing, learners' response to reformulation was positive, note-taking as a measure of noticing elicited valuable data, Hanaoka's (2007) measuring procedure was viable and the questionnaire added support to other data.

In addition to the qualitative (noticing-sheets and questionnaire's open-ended questions) and quantitative (essay error analysis and questionnaire's multiple-choice questions) results, testing the design, data collection methods, framework of analysis and identifying unforeseen problems were the most important outcomes from the pilot study. The implications for the main study were various. Below I first look at design features that were retained in the main study. Then, I identify findings that reshaped the main design.

6.1 Retained features

The writing task (genre, topic, instructions) was retained. Open tasks generated authentic language use. The topic stimulated writing and allowed learners to write based on their everyday knowledge.

Note-taking also encouraged learner-initiated noticing in the composing and feedback stages in a non-disturbing way. The freedom and autonomy given to learners elicited much valuable information. Note-taking revealed individual differences regarding introspection, feedback processing and self-awareness skills. Therefore, noticing sheet-1 and noticing sheet-2 were retained as the key measure of noticing. Brief, simple instructions on noticing sheet-1 asking learners to report specific problems rather than general language areas worked well. Inclusion of noticing sheet-

1 in the design was also a good decision as, consistent with the questionnaire results and NS1-NS2 comparison, it predisposed learners to attend to their feedback.

Regarding 'reformulation', learners' positive response to this feedback technique made me eager to explore this technique further. Its effectiveness for noticing was reflected in the noticing sheets. Reformulation gave learners self-confidence as they noticed both their errors and their strengths. Nobody experienced text appropriation and learners noticed cultural features in their essay writing. Reformulation also encouraged learners' meta-reflection on their noticing. Such reflections might be evidence of different levels of awareness, e.g. "I used to think that if I did not know a word I could just define it. But I now realize that if I do that, the text loses coherence and it may even be misunderstood" (Student-1). Learners' questions about changes they did not understand were also evidence of L2 processing, e.g. "I do not understand the use of very in 'They bottle the very water they take from…', I think it adds emphasis, I had never seen this use of very" (Student-4). Learners also differentiated reformulation (they called it authentic feedback) from mere correction (they called it error correction). They stated error correction made them believe their L2 problems were reduced to marked errors, whereas reformulation allowed them to see other writing requirements they need to become more native-like.

Essay error frequency analysis provided quantitative evidence on the learners' major grammar problems. Learners' top grammar type errors were in syntax: unnecessary definite article (UDA), omission of subject, subject + verb agreement, prepositions, verb + preposition/ no preposition, word order, subject + possessive adjective agreement. These results provided quantitative evidence for structures only inferred by the CELE teachers. Error rate analysis was repeated with a larger sample in the main study.

The small sample (ten students) allowed me to try Hanaoka's (2007) measuring procedures and categories (problematic features noticed, PFNs; features noticed, FNs). Hanaoka's problematic features noticed and features noticed were retained, but his measuring procedures were not feasible in the main study. Noticing sheets analyses in the pilot study was qualitative, I read and analysed the data on noticing sheets, identified categories and added a tick every time a learner reported a feature under a category. Categories were created and modified based on emerging data. The results showed that lexis and word repetition were the features most noticed; L1 translation and punctuation were noticed equally; connectors came next. Prepositions and spelling came last and were identified in the same proportions. See Table 8 to compare the results for essays and noticing sheet-2.

Table 8 – *Learners' reported FN and top recurring grammatical errors in the essays.* From top to bottom (from most to least frequent).

Top recurring grammatical errors (essays error frequency analysis)	**Learners' reported features noticed in reformulated essays (NS-2)**
Unnecessary definite article (syntax)	Lexis (non-grammatical)
Omission of subject (syntax)	Word repetition (non-grammatical)
Subject + verb agreement (syntax)	L1 translation (grammatical-syntax)
Prepositions (syntax)	Punctuation (non-grammatical)
Verb + prep. / no prep. (syntax)	Connectors (non-grammatical)
Word order (syntax)	Prepositions (grammatical-syntax)
Subject + poss. adj. agreement (syntax)	Spelling (non-grammatical)

Note. FN = features noticed, *NS* = noticing sheet, *prep.* = preposition,
Poss. adj. = possessive adjective, *L1* = mother tongue
Essay analyses included only grammatical features.
Noticing sheet-2 (NS-2) analyses included grammatical and non-grammatical features.

The number of features noticed per student was also calculated. Interestingly, Student-1 (whose essay was reasonably well written) noticed the largest number of features (8), whereas student-10 (whose essay was poorly written) reported the least number of features noticed (1). This suggests that learners' amount of noticing might be related to learners' L2 proficiency.

Concerning the exit questionnaire, the data obtained expanded and reinforced observations on the noticing sheets. A revised, shorter version was preferred for the main research.

6.2 Redefining the design

Reducing the text length (300 words) and strict adherence to instructions were lessons learned from the pilot study. Including a checklist, administering learners' time on-task (indicating when composing time ends and proofreading time starts) became possibilities to consider.

Abandoning the original idea of comparing two types of noticing sheets in the feedback stage had positive outcomes. Originally, learner-initiated and teacher-prompted noticing sheets were to be compared. However, using the same learner-initiated noticing sheet for all participants was successful. The implemented noticing sheet-2 thrived in promoting learner-initiated noticing and processing of feedback. Learners' reports revealed what was important for them in their feedback, why and how well they understood (if they did) their corrections.

Implementation of the pilot study itself made me aware of decisive features for the main design. A whole class talk with the students was held after the pilot experience. In that session, students received their direct error corrected essays as corrections had by then been made by the researcher for error analysis. It was considered worthwhile for the students to have their corrected essays. During this procedure, I observed learners' negative reaction towards the direct error correction technique, especially after trying reformulation. I wondered what the impact on noticing would be if the two WCF techniques were compared. This thought became a reality in the new research design. Maintaining my interest in learner-initiated noticing, I decided to compare the types of noticing that result from two different

types of WCF (reformulation and direct error correction). The pilot experience made me aware this might be more relevant for research, learners and L2 teachers. Furthermore, if one experimental rather than all experimental groups used reformulation the amount of work for the reformulators would decrease.

The mismatch between learners' grammar problems in essay writing and their reported features noticed (Table 8) was a striking finding. The results were expected to differ as error analysis exclusively considered grammar components, whereas learners' features noticed included all types of errors: lexis, syntax, grammar, discourse, punctuation features. Yet, as Table 8 shows, learners' noticing in reformulated essays hardly dealt with grammar. Prepositions, the only grammar feature noticed (L1 translation was considered a broad category requiring further characterisation), came sixth, after lexis and other discourse features, such as word repetition or connectors. The impact of reformulation on learners' noticing was observed in non-grammatical (lexis, discourse, punctuation) rather than grammatical features. Would direct error correction have the same effect on learners' noticing? This was another question that influenced the new research design.

The aforementioned mismatch supports Van Patten's (1994) claim that, in language processing, content rather than linguistic features is processed first, accuracy is secondary for comprehension – and for WCF via reformulation I would add too.

Learners' features noticed reported on noticing sheet-2 corresponded with reformulators' perceived problems in learners' essays. Both reformulators and learners identified lexis, word repetition, punctuation, connectors, spelling and L1 translation. Learners' features noticed also included prepositions and errors that were particularly theirs. What is significant in this correspondence is the fact that learners' features noticed derived from learner-initiated noticing. Students themselves became

aware of these features and went through an awareness experience. Reformulators' perceptions were, on the other hand, mere estimations with no quantitative support before the error frequency analysis. Finally, it was advised that the pen-and-paper data collection process be computed-supported to accommodate students', reformulators' and the researcher's work.

SECTION 3: TOWARDS THE MAIN STUDY

Chapter 7: Complexity of Comprehensive Error Correction

In Chapter 2, I explained that focused (or non-comprehensive) WCF is a type of error correction (EC) that addresses specific linguistic targets. It is usually contrasted with unfocused (or comprehensive) WCF, a type of error correction that deals with all types of errors (syntax, morphology, lexis, style, cohesion, punctuation etc.). Although focused or non-comprehensive error correction is not a technique representative of L2 classrooms, it is the prevailing one in WCF research. Why is this so? Van Beuningen (2011) explains that focused EC prevails because it is easier for research. However, she delineates its constraints for research too: consideration of a limited number of errors; learners' tendency to monitor their use of the target structure; writing tasks resembling grammar exercises; an unauthentic correction technique. Chapter 2 also conducted a discussion on the advantages and disadvantages of unfocused WCF. Theoretical and practical arguments 'for' and 'against' the use of unfocused or comprehensive WCF were mentioned. Table 9 briefly summarizes those arguments.

Based on a whole-class conversation with students (pilot study) and discussions with teacher-raters in training sessions (main study), I confirmed Lee's (2005) findings about teachers' and students' preference for comprehensive error correction being applicable to the CELE context. As Mantello (1997) points out, despite its disadvantages and the wide variety in teachers' correction practices, "what tends to remain a constant … is the notion that we must address all errors occurring in students' writing. We must correct errors comprehensively" (p. 128). On the other hand, studies that show evidence to support the efficacy of WCF have used focused WCF. Generalization of their results is though impossible, as their evidence is restricted to a limited number of structures (Storch, 2010).

Given my aim of performing research that resembles EFL classrooms as closely as possible, and accounting for the limitations of focused WCF and considering Mantello's (1997) and Lee's (2005) observations, I decided to focus my study on direct comprehensive EC. Research in comprehensive EC is scarce. It is assumed to be difficult for research yet its difficulties are not really explained.

7.1 Emerging problems

Despite the warnings against performing research on comprehensive EC, the few existing studies on this correction type (Ellis et al., 2008; Sheen et al., 2009; Truscott & Hsu, 2008; Van Beuningen, 2008, 2011) encouraged me to move forwards. Hence, for almost two months, I corrected 135 300-word essays. Correction was comprehensive and included the correction of non-grammatical (lexis, spelling, connectors, punctuation, pragmatics) and grammatical (syntax, morphology) errors. The error categorization chart constructed for the pilot study was reused and improved for correction in the main study. Non-grammatical categories were added (as the pilot study showed non-grammatical features were frequently reported as noticed) and grammatical categories were better defined. See the final error categorization in Appendices A and B for grammatical and non-grammatical errors, respectively. This meticulous error correction familiarized me with the linguistic data in the learners' essays. I intuitively perceived not only learners' most frequent errors but also initial trends between different groups' writing performances.

Table 9 – *For and against arguments for comprehensive or unfocused EC.*

Comprehensive or unfocused error correction		
Argument	**Against**	**For**
Theoretical	-Krashen's Natural Order Hypothesis. -Pienemann's Teachability Hypothesis. -Schmidt's view on 'attention' being limited.	Constraints of focused error correction for research (Van Beuningen, 2011): -Focus on a limited number of errors (usually uncomplicated ones for research). -Learners' tendency to monitor their use of the target structure. -Writing tasks resembling grammar exercises. -Unauthentic error correction technique.
Practical	Comprehensive WCF is: -confusing -inconsistent -unsystematic -overwhelming … for students and teachers!	Evans et al. (2010): -Demands for L2 accuracy and precision from academic and professional spheres. -Teachers' ethical obligations. -No experimental evidence on the inefficiency of comprehensive WCF.

Note. *EC* = Error correction; *WCF* = Written corrective feedback

When 75 per cent (135) of the total essay sample (180) was corrected, I invited three CELE English teachers to help me with inter-reliability analyses. Common features among these teachers were Spanish speakers, a minimum of ten years' teacher experience, holding a bachelor's degree (linguistics, chemistry, dentistry) and a CELE teaching diploma. The teachers participated in training sessions whose objectives were to prepare them in the use of the error categorization devised by the researcher and test its efficacy. The teacher-raters and the researcher corrected one essay together. Then each teacher-rater (four including the researcher) corrected three essays individually. Individual corrections were compared in consecutive sessions where the numbers of errors in each error category were confronted and their categorizations justified. Discrepancies were expected to be resolved through

discussion. However, the task immediately revealed: different teachers' correction practices (some teachers tend to over mark while others tend to be more lenient; some teachers tend to reconstruct, while others tend to reformulate); different views of writing and error correction; frequent undetected errors by teacher-raters (the researcher included). These dissimilarities in teacher-raters' error correction practices might have been "random and arbitrary", as with Zamel's teachers in her 1985 study. More frequently, however, participant teacher-raters' approaches to errors had been learned in teacher training and development courses, and tested in their teaching practice. Concerning undetected errors in this particular experience, these were not due to raters' inability to correct them, but to the different physical, mental and environmental conditions in which each rater reported s/he had corrected; the overwhelming nature of comprehensive EC is likely to have played a role too.

None of the above problems was as difficult to deal with as the error categorization of certain linguistic features. Most parts of speech are multi-functional. A 'verb + preposition' error, for instance, can be categorised as a lexical or a preposition error. Similarly, the word 'healty' in the sentence 'Even if it is dangerous for their healty' could be categorized as a 'wrong word error', an 'inflection error', a 'lexical error' or a spelling error. In other cases, e.g. 'TV imposes beauty stereotypes teenagers adopt without questioning', a rater would fix specific learners' faults by inserting a full stop and starting a new sentence: 'TV imposes beauty stereotypes. Teenagers adopt them without questioning'; but another rater would correct the same error by linking two ideas with a relative pronoun: 'TV imposes beauty stereotypes that teenagers adopt without questioning'; a third rater would see the sentence as correct. All raters' responses above are correct; nevertheless, they led to differences in the total error count. Raters' discussions and agreements continued over ten two-hour

weekly sessions. By then I realized that comprehensive EC was extremely subjective and a discussion of 180 essays was unfeasible.

7.2 From comprehensive to semi-comprehensive EC

Ferris (2011) says that arguments and counterarguments about comprehensive EC are real and deserve attention. Although I was familiar with both views, my inclination for comprehensive EC determined my research design. The problems encountered during implementation made me (a) question its practicality for experimental research, (b) aware of the impact of people's beliefs on everyday practices and, more importantly, (c) understand that my predisposition towards comprehensive EC derived not only from my desire to investigate what actually happens in the classroom, but also from my belief in comprehensive error correction. Lee's (2004) findings about teachers' and students' preferences for comprehensive EC do not imply that such preferences are relevant to writing development. Zamel (1985) adds that teachers' beliefs come from "experiences, orientations, expectations, preconceptions, and biases" (p. 82). In other words, what is common is not necessarily what is certain. Therefore, what Lee's study shows is that teachers' correction beliefs (based on principles or bias) are transmitted to students, who accept them as correct, in everyday instruction.

After tackling the challenges of comprehensive EC, the solution I found to move forward with my study was to opt for semi-comprehensive error correction, i.e. the selection of seven linguistic features to focus on. Whether learners noticed the selected errors (or not) became part of my research.

Chapter 8: Data Analysis and Results

8.1 Analysis and results

Table 10 shows an average of 50 (grammatical and non-grammatical) errors per 300 words resulting from comprehensive error correction of essay 1 in the four participant groups (60 essays in total, 15 essays per group): DIR 49.4; REF 49.9; SELF 57.6; CONTROL 43.7. Which and how many linguistic features to target were decided quantitatively and qualitatively.

Table 10 – *Analysis of essay 1 in the four groups.*
Numbers of errors, error average and standard deviations between groups.

	DIR	**REF**	**SELF**	**CONTROL**
Error average	49.4	49.9	57.6	43.7
Standard deviation	11.5	13.1	11.7	16.2
Maximum	74	67	72	74
Minimum	34	28	26	13
Total no. of errors per group	741	749	864	656
Total no. of errors in 60 essays			3,010	

Note. DIR = Direct group, *REF* = Reformulation group, *SELF* = Self-correction group

An error frequency analysis (like the one in the pilot study) of essay 1 in the four participant groups was conducted. Appendix C shows the most frequent error types. This analysis gave quantitative support to error types only intuitively perceived by teacher-raters. Appendix C corresponds to the complete error frequency analysis.

Table 11, below, shows only the most frequent errors: 'spelling' (467) and 'lexis' (371) in the non-grammatical error type. In the syntax category of grammatical error type, 'wrong word' (240) came first, with almost the same frequency as 'omission of constituent'[12] (239), 'unnecessary definite article UDA' came second (148) and 'verb tenses' (77) was third, but is not included in Table 11 as it is a

[12] A constituent is a linguistic unit. Constituency is the relationship between a constituent and the larger unit that it is a part of. A constituent can be a morpheme, word, phrase or clause.

category that includes various tenses. ‘S+V agreement’ (62) and ‘V + preposition/ no preposition’ (62) had the same frequency. Regarding the morphology category of grammatical type, ‘gerunds’ (74) was the most frequent error, closely followed by ‘singular for plural’ (71) in second place and ‘verb form’ (69) third, closely followed by ‘third person singular-3PS’ (63) in fourth place.

Given the difficulties of comprehensive EC (highly subjective, multi-functional nature of linguistic items, different ways of approaching the same errors), the teacher-raters and the researcher concentrated first on the most frequent errors. Among them, we selected clear-cut unambiguous errors whose identification was as unequivocal as possible. Each of the most frequent features in Table 11 was discussed and classified as ambiguous or unambiguous for teacher-raters; then, clear criteria for its categorization were defined. Selected non-grammatical errors included ‘lexis’ and ‘spelling’. Selected grammatical errors were ‘omission of constituent’, ‘unnecessary definite article’ and ‘subject + verb agreement’ for syntax, ‘gerunds’ and ‘3rd person singular’ for morphology. ‘Wrong word’ resulted in one of the most ambiguous errors frequently overlapping with lexis. Thus, we decided not to select it. As for ‘unnecessary definite article’ and ‘subject + verb agreement’, they were selected for their objectivity. Concerning morphology, we found ‘gerunds’ and ‘3rd person singular’ easier to identify than ‘verb form’, and more important for our criteria than ‘singular for plural’. Seven error types were selected. Ferris (2010) observes that nobody knows “what the most appropriate number of written CF categories should be” (p. 196). The ‘Error frequency grid’ used by the raters and researcher in semi-comprehensive EC is shown in Appendix E. The one used by the researcher in the original attempt to deal with comprehensive error correction is in Appendix D.

Table 11 – *Top errors in error frequency analysis of essay 1 (four groups).*

Type of error & error category	Structure	DIR (1–15)	REF (16–30)	SELF (31–45)	CON (46–60)	Total	%
NON-GRAMMATICAL							
Spelling		124	103	142	98	467	15.5
Lexis		122	86	98	65	371	12.3
GRAMMATICAL							
Syntactical	UDA	40	40	43	25	148	4.9
	S+V agree.	23	15	15	9	62	2.1
	Wrong word	31	49	89	71	240	8.0
	Omission of constituent	38	53	85	63	239	7.9
	Verb tenses	22	17	18	20	77	2.6
	Verb + prep.	22	12	14	14	62	2.1
Morphology	3PS	15	21	12	15	63	2.1
	Verb form	18	22	17	12	69	2.3
	Sing. for plural	8	15	24	24	71	2.4
	Gerunds	15	22	17	20	74	2.5

Note. *DIR* = Direct group, *REF* = Reformulation group, *SELF* = Self-correction group, *CON* = Control group, *UDA* = Unnecessary definite article, *S+V agree*. = subject + verb agreement, *prep.* = preposition, *3PS* = 3rd person singular

8.2 Criteria to describe selected error categories

Together, the teacher-raters and the researcher set criteria for the categorization of each of the selected errors. Agreement on certain error ambiguities was reached.

Omission of constituent: This category includes the omission of verbs, nouns, prepositions, articles, personal pronouns and subject, e.g. verbs: ‘The story ___ about a man who’; nouns: ‘Their way of talking, their way of dressing and their body ___’; prepositions: ‘Even models suffer ___ this effect’; articles: ‘This creates a mismatch between the meaning of ___ perfect body and beauty’; omission of subject: ‘In the end ___ is just a lie’. We perceived that the ‘omission of constituent(s)’ increased when raters were trying to reformulate rather than reconstruct students’ writing. Therefore, respect for learners’ original writing was given priority. Raters avoided modifying learners’ original sentences. It was also noted that although

‘omission of constituent’ may include more than one word, the addition of more than three words was a signal of rater’s reformulation rather than reconstruction. ‘Omission of auxiliary verbs’ was not included in this category but rather in the tense category, e.g. ‘Now, we ___ going to talk about’.

Unnecessary definite article (UDA): ‘The’ is an article used to refer to a specific thing, quantity or group. Focus was given to the most common problem for Spanish speakers, i.e. the overuse of ‘the’. We corrected all errors where ‘the’ did not refer to a specific thing, quantity or group.

Subject + verb agreement: The basic rule is: a singular subject (The box of nails) takes a singular verb (is broken), whereas a plural subject (The nails) takes a plural verb (are old). However, Spanish speakers frequently fail to use S+V agreement correctly because they cannot identify the main noun in compound subjects or because of English collective nouns, such as ‘people’, ‘everybody’, children etc. 3PS can also be classified as S+V agreement errors. However, we agreed that unless that 3PS referred to have/has or be/is, it would be classified as a 3PS error.

Third person singular: This category was reduced to morphological errors, i.e. the addition of ‘s’ or ‘es’ to verbs in the simple present. We also agreed that the verb ‘to be’ would not be included in this category. Therefore, the following examples were classified as ‘S+V agreement’ errors: ‘This problem have existed’, ‘New phenomena emerge in society, one of the most important are the cultural industry.’ The addition of ‘s’ to the main verb when the auxiliary ‘does’ already exists was not included in this category either. Thus, ‘Does it really matters?’ was considered a ‘verb form’ error.

Gerunds: This category will be reduced to morphological errors, i.e. the addition or omission of 'ing' to verbs. Gerunds include functions such as nouns, adjectives and 'ing' as infinitives. Progressive tenses were not included.

Spelling: Differences in error identification were mainly due to hard-to-perceive spelling mistakes. This problem was solved by using the spell-check function in a word processor.

Lexis: This category deserved special attention, as it resulted in an extremely ambiguous category that overlapped with 'wrong word, inflections, V + preposition' and even with 'connectors'. However, as it was learners' second most frequent error, it had to be included. Criteria were defined to recognise this error category. Criteria took into consideration the way learners look at their errors. Perhaps, because of learners' meagre metalanguage, they tend to simplify wrong words, inflections and even connectors due to lack of vocabulary. In other cases, learners' errors are better explained and understood from a lexical point of view. Below are emerging lexis problems and agreements.

Problem 1: Learners' lexis is, most of the time, understandable though not always the best choice, for example: In the 20th century, new concepts 'arrived to' society (emerged in, appeared in); The cultural industry 'creates a fight' between the meanings of two words (leads to disagreement, causes mismatch).

Agreement: Keeping in mind that upper-intermediate learners' main goal is to move beyond the plateau on which they are able to communicate though not always in a native-like way, raters agreed on correcting everything that, despite being understandable, English speakers do not usually say. The view of writing as an L2 production opportunity in which learners use their own means to communicate and later receive feedback as pushed output also guided our decision. Therefore, lexis was

not reduced to words hindering communication but included the correction of everything that was not native-like.

Problem 2: 'Verb + preposition/no preposition'. Verbs followed by their corresponding preposition were considered lexical problems, e.g. 'Today's accepted bodies are different that (different from) bodies accepted in the sixties.'

Agreement: Bearing in mind the way learners would have their errors explained, the raters agreed that lexis should account for lexical chunks rather than isolated verbs and prepositions. Many English expressions can be explained as lexical chunks. These include "verbs/adjectives + preposition", e.g. worried about, jealous of.

Problem 3: Lexis also overlaps with other categories, such as connectors, wrong word, inflections and word families.

Considering that most learners see these categories as a lack of vocabulary, they were classified as lexis, e.g. 'Nowadays the improve of cameras…' (improvement/ inflection).

Agreement: Due to learners' meagre metalanguage, inflections might be better understood as word families, e.g. 'Beautifulness and money are over-valued in today's society' (word family: beauty, beautiful, beautifully); 'What the saying says is complete true' (completely/ inflection; but also word family: complete, incomplete, completely).

Problem 4: Grammar vs. lexical errors? Quite often, errors that might initially be perceived as grammatical errors actually have lexical causes, e.g. It was then that the 'mini-skirts' borned (were born).

Agreement: The above sentence reflects the learner's partial knowledge of verbs, a problem that might be overcome if 'to be born' is explained to him/her as a

compound verb or lexical chunk. Therefore, whenever an error can be explained as a lexical chunk this will be preferred.

Problem 5: 'Collocation' can also be considered a wrong word error, e.g. 'do research' vs. 'make research'?

Agreement: All collocation problems will be considered lexical errors whether they are wrong words or not.

Problem 6: No error will be corrected twice, e.g. 'Even if it is dangerous for their healty' (healthy/health). Whether this is a 'wrong word', 'inflection' or 'lexical' error, it is first a spelling mistake, should it be corrected twice?

Agreement: No error will be corrected twice. The type of correction that best solves the error will be preferred. In the above example, 'healty' is first a spelling error. However, if only the spelling is modified the sentence will still be incorrect and the learner will not know his lexis choice was incorrect. Thus, this should be classified as a lexical rather than a spelling error.

8.2.1 Other lexis agreements

False cognates: All false cognates were considered lexical errors, e.g. 'Before becoming famous, those people were not important actually they are canons of beauty' (At present, today).

'Be like' vs. 'look like': The essay topic in this writing task made the use of 'be like' and 'look like' quite frequent. This frequency showed the distinction between these expressions was not clear for learners. Despite the use of both expressions being grammatically correct, there were contexts where one expression described more accurately what students meant to say, e.g. 'Girls that grow up with the idea of skinny bodies as perfect bodies want to be like their models' (to look like). The frequency grid used in semi-comprehensive EC can be found in Appendix E.

8.3 Testing reliability

Five per cent (twelve) of the 180-essay sample went through intra- and inter-reliability analysis. Three of the essays were corrected and discussed during training sessions. The remaining nine essays were corrected individually, focusing exclusively on the selected features and using the agreed criteria for each error category. I made sure essay correction was blind and the sample was adequately randomized.

Following Van Beuningen's (2011) study, I calculated intra-class correlation coefficients (ICC) to find the average rankings of inter-rater and intra-rater agreement for overall, non-grammatical and grammatical accuracy (Table 12). ICC estimates the correlations between individual measurements and between average measurements made of the same target. A STATA two-way random effect model was used to estimate the inter-rater ICCs. The researcher corrected the same nine essays at two different moments (first analysis: January 2015; second analysis: June 2015) for the intra-rater ICCs. As observed in Table 12, the result for the ICC for inter-raters was lower for non-grammatical accuracy. This is explained by the difficulty of lexis categorizations.

Table 12 – *Intra-class correlation coefficient (ICC).*

	Grammatical accuracy	Non-grammatical accuracy	Overall accuracy
ICC inter-rater	.80	.74	.87
ICC intra-rater	.93	.84	.90

8.4 Insights from coping with comprehensive EC

In this chapter, I have described the difficulties I encountered when performing direct comprehensive error correction on open uncontrolled 300-word essays. Despite my literature review of previous studies, the full complexity of comprehensive EC was not foreseen during the design. Van Beuningen's (2011)

comprehensive EC dealt with two broad categories: grammatical and non-grammatical. Follow-up of more specific structures was only performed for four students. Ellis et al. (2008), on the other hand, called comprehensive EC what I have called 'semi-comprehensive error correction', i.e. they also focused on a selection of errors. The inclusion of more than two or three grammar structures (as most WCF studies do) might be the reason why Van Beuningen et al. and Ellis et al. call their techniques comprehensive. Although I did not aim to provide evidence for the efficiency or inefficiency of direct comprehensive EC, I have described the problems encountered and the criteria used for error categorization and selection of the seven target features used in the essay analysis. The complexity of the technique might be a result of the open, uncontrolled nature of the writing task used in this study. Different from controlled WCF studies, in opinion essays there is no single way to do the task, and no single model to contrast what is correct or incorrect. Crucial for this study on learner-initiated noticing was the absence of pre-determined linguistic targets, which originally justified the use of comprehensive error correction. Selection of error categories was necessary for methodological reasons. Making error correction semi-comprehensive made essay analysis feasible and allowed me to contrast learners' linguistic needs with what they pay attention to in their feedback.

The selection of errors for this study was based on a systematic analysis that provided quantitative information about our particular learners' errors. These data might be useful to guide CELE EFL instruction, syllabus and materials design. Raised awareness about different error correction practices and beliefs was an important benefit for both the teacher-raters and the researcher. The raters and I realized that there were different ways of correcting the same ideas and that we responded differently to the same text. There exist different ways of understanding "error", of

distinguishing “errors” from “improved changes” and of deciding what is acceptable and what is not. ‘Error gravity’, Giraldo de Londoño and Perry (2008) say, “is determined not only by the kind of error committed but also by its frequency, the circumstances in which it is committed, the identity of the one committing it, and the type of person(s) judging it” (p. 113). The notions of ‘reformulation’ and ‘reconstruction’ were constantly present in our discussions. Some raters realized they tend to reconstruct, while others tend to reformulate. ‘Text appropriation’ was also frequent. Raters repeatedly alerted each other when anyone’s correction was changing learners’ original ideas. Eventually, as raters, teachers and researchers, we wondered and worried about the effects of all our discrepancies on students’ learning. After all, we were only four raters correcting the same three essays. The differences in our error correction practices were just a sample of classroom reality. As Zamel concluded in her 1985 study, students are “likely to be confused by the contradictory ways in which different teachers respond” (p. 82). Cumming (1983) explains that teachers’ obsession with errors originates from their own perception as language teachers who deal with the prescribed aspects of language only. Teachers have failed to see their role as teachers of writing in charge of learners’ cognitive development.

SECTION 4: MAIN STUDY

This section comprises Chapters 9, 10 and 11. Chapter 9 states the research questions (RQs); describes the participants and setting; presents the data collection methods; explains the design, treatments and experimental procedures. The last section in this chapter explains some methodological manipulations or strategies in the design. Chapter 10 describes the analyses of data in the essays, noticing sheets and questionnaires. Chapter 11 reports the results obtained from the essays, noticing sheets and exit questionnaires.

Previous studies strongly influenced my design. From Hanaoka (2007), Santos et al. (2010) and Park (2011) I share learners' reports in the form of note-taking as a way to collect evidence of noticing. From Hanaoka (2007) and Hanaoka and Izumi (2012), I follow their suggestion to include a noticing sheet in the composing stage. With Swain (1985) and Qi and Lapkin (2001), I share their conviction that noticing takes place in both input and output. Van Beuningen's (2011) research greatly shaped my design after the pilot study. Her study encouraged me to categorize errors as grammatical vs. non-grammatical types and to give my study a quantitative direction. Learners' accuracy performance in pre-, post- and delayed post-essays was compared. Accuracy is important in the writing-to-learn dimension, as learners' L2 writing ability is mostly assessed for its accuracy. Accuracy concerns both grammatical and non-grammatical features.

Chapter 9: Method

9.1 Research Questions

These are the research questions (RQs) that guided my study:

RQ-1: What are the effects (if any) of different comprehensive (eventually replaced by semi-comprehensive) error correction conditions (direct, reformulation and self) on learner-initiated noticing at two stages (composing and feedback) of the writing process?

1.a - Do different comprehensive (eventually replaced by semi-comprehensive) error correction conditions (direct, reformulation and self) lead to learner-initiated noticing in the composing and feedback stages? i.e. The occurrence of noticing reports.

1.b - If so, how much noticing and what error types (grammatical, non-grammatical) do learners pay attention to by themselves in each error correction condition at the composing and feedback stages, i.e. The number of noticing reports and error types reported as noticed.

RQ-2: What are the effects (if any) of the above noticing-supported error correction conditions at the feedback stage on learners' writing accuracy in the rewriting and new writing stages?

2.a - Does the input provided by the above noticing-supported error correction conditions at the feedback stage have any effects on learners' writing accuracy in the rewriting and new writing stages? i.e. The occurrence of effects of noticing-supported error correction conditions in the rewriting and new writing stages.

2.b - If so, what error types (grammatical or non-grammatical) are more amenable to correction in different noticing-supported error correction conditions? i.e. The error types learners corrected by themselves in different error correction conditions.

9.2 Participants and setting

This study was conducted at the Foreign Language Teaching Centre (CELE) of the National Autonomous University of Mexico (UNAM) with students in tertiary education. Despite English proficiency constitutes a graduation requirement for UNAM students, neither credits are gained nor requirements are met by studying at CELE. English becomes an extracurricular subject and students attend courses on a voluntary basis. English as a foreign language is taught at CELE with a communicative approach in four-skill courses. Classes were two hours long and took place three times per week.

The population of this study consisted of 60 students (*N = 60*) in their second year of tertiary education. Participants were divided into four groups of 15 students each (*N = 15*). All participants were born in Mexico and were monolingual Spanish speakers who started learning English at the age of 13 (as English is taught compulsory in Mexican secondary schools). Participants' English level was upper-intermediate (equivalent to B2 level in the Common European Frame of Reference). Participants' demographic profile is presented in Table 13 where participant groups are listed with the ECC that was assigned to each one of them: DIR-EC for group 1, REF-EC for group 2, SELF-EC for group 3 and, CTRL for group 4.

Table 13 – *Demographic profile of the sample.*

Feature	**Group**	**PER GROUP** *(N = 15)*						**TOTAL SAMPLE** (*N = 60*)					
		mean	***SD***	**min-max**				**mean**	***SD***	**min-max**			
Age	DIR	26.5	6.3	20 – 38				23.6	4	19 – 38			
	REF	22.4	1.9	19 – 26									
	SELF	21.7	0.9	21 – 23									
	CTRL	23.9	2.3	21 – 29									
Gender	**Group**	**males**	**females**					**males**	**females**				
	DIR	8	7					27 (45%)	33 (55%)				
	REF	6	9										
	SELF	2	13										
	CTRL	11	4										
Major	**Group**	SCI	SOC	A&H	BA	A	NR	**SCI**	SOC	A&H	BA	A	NR
	DIR	5	4	4	0	0	2	46%	23.3%	11.6%	10%	3.3%	5%
	REF	6	5	1	1	1	1						
	SELF	5	4	0	5	1	0						
	CTRL	12	1	2	0	0	0						

Note. *DIR = Direct group, REF = Reformulation group, SELF = Self-correction group, CTRL = control group, SD = Standard Deviation, min. = minimum, max. = maximum, SCI = Scientific Sciences, SOC = Social Sciences, A&H = Arts and Humanities, BA = Business Administration, A = Accounting, NR = Not reported.*

Table 13 shows that the mean age of participants' total sample was 23.6, *SD* = 4 (min. 19, max. 38). Mean age per group was: DIR: 26.5, *SD* = 6.3 (min. 20, max. 38); REF: 22.4, *SD* = 1.9 (min. 19, max. 26); SELF: 21.7, *SD* = 0.9 (min. 21, max. 23); CTRL: 23.9, *SD* = 2.3 (min. 21, max. 29).

Concerning gender, Table 13 shows the number of females (33 or 55%) exceeded the number of males (27 or 45%) in the total sample. Gender data per group were: DIR: 8 males, 7 females; REF: 6 males, 9 females; SELF: 2 males, 13 females; CNTRL: 11 males, 4 females.

The total sample was heterogeneous with respect to participants' majors too. Scientific Sciences (e.g. Engineering, Chemistry, Physics) predominated (46%, i.e. 28 out of 60 participants); Social Sciences (e.g. Communications, Psychology, Sociology) were second (23.3%, i.e. 14 out of 60); Art and Humanities (Literature, Education) were third (11.6%, i.e. 7 out of 60); Business Administration was fourth (10%, i.e. 6 out of 60), followed by Accounting (3.3%, i.e. 2 out of 60) and majors not reported (5%, i.e.3). Majors per group were: DIR: Scientific Sciences (5), Social Sciences (4), Art and Humanities (4) and majors not reported (2); REF: Scientific Sciences (6), Social Sciences (5), Art and Humanities (1), Business (1), Accounting (1) and major not reported (1); SELF: Scientific Sciences (5), Social Sciences (4), Business (5), Accounting (1); CNTRL: Scientific Sciences (12), Social Sciences (1), Art and Humanities (2).

Participants were recruited as they registered for their upper-intermediate EFL term. Experimental and control conditions were randomly assigned to the four groups. The four groups participated in the same four-stage writing task (composing/ error correction-noticing/ rewriting/ new writing) and experienced the same treatment. ECCs were different for each group. All tasks and treatments were administered

during class periods. The researcher introduced and explained the tasks. Class teachers were absent from all experimental sessions. Students' participation was both anonymous and voluntary.

9.3 Data collection methods

Learners' written essays, two noticing sheets and an exit questionnaire constituted the data collection methods. They were all tested in a pilot study and improved based on its results. This section describes the final versions of these methods and the rationale for their design.

9.3.1 Written essays

Written essays aimed to collect evidence of learners' writing accuracy in the: composing (essay 1), rewriting (essay 2) and new writing (essay 3) stages. The writing task was an open, uncontrolled, learner-centred, 300-word opinion essay. The writing topic (tested in a pilot study) was found to be appealing, provoking and familiar to the CELE student population. A familiar topic was necessary if learners were to write about it based on their experience and everyday knowledge. The ability to write opinion essays is a syllabus-based task for the target students. Conformity with the school curriculum was essential to satisfy learners' expectations and ensure administrative and teachers' support. However, the writing task implemented was exclusively designed for this study and learners did not receive any grade for their participation. The word limit was carefully combined with other task demands, such as planning, writing, proofreading and word count, all tested in a pilot study. Prompts for the new task demanded much careful planning and testing before their implementation. See essay final versions in appendices H (WT-1), I (WT-2) and J (WT-3).

9.3.2 Noticing sheets

Two noticing sheets were used to encourage learners' noticing at two different stages of the writing process: (a) 'Noticing the hole' while writing (NS-1) and (b) 'noticing the gap' between learners' developing linguistic system and the L2 system while looking at their feedback (NS-2). NS-1 kept a record of learners' reported noticed limitations while composing, whereas NS-2 kept a record of learners' reported noticed errors while receiving feedback. I opted for NSs due to the advantages they offer for research: (a) they capture attentional processes in real time, in a non-disturbing way, and are reasonably easily compared with other methods; (b) they allow for learners' autonomous analysis of input; (c) learners' notes specify where learners concentrate their attention and provide evidence of learners' nature of awareness (Hanaoka, 2007), i.e. whether a noticed feature is new or familiar to the learner; (d) they have been used successfully in previous studies dealing with noticing in WCF (Hanaoka, 2007; Hanaoka and Izumi, 2012; Santos et al., 2010); (e) the results from the pilot study for this research showed their potential to provoke and provide information about learners' internal cognitive processes: noticing, processing of feedback, writing strategies and attention levels, to mention some.

In Chapter 4, I explained that learner-initiated noticing entails responding to noticing opportunities being a learner's decision. Learners decide whether and to what they pay attention. With this in mind, the instructions on NS-1 and NS-2 gave learners the option to leave the sheet blank if they wanted to (discussed further in Chapter 10). This was to guarantee that noticing was voluntary and that task instructions did not force learners to report anything. Self-provided or self-initiated feedback made this offer unfeasible for the self-correction group.

The instructions on NS-1 (Appendix F) were the same for all three experimental groups. The instructions on NS-2 (Appendix G) were the same only for the direct (DIR) and reformulation (REF) groups. A comparison of an original text with a 'corrected' or 'reformulated' essay, as the instructions read, was a small but necessary change. Self-provided or self-initiated feedback in the self-correction group called for different instructions (see appendix G).

Noticing sheet 1: NS-1 was used in the composing stage to encourage output noticing. Acknowledging Hanaoka and Izumi's (2012) negative experience with detailed long instructions in similar NSs, the instructions on NS-1 in this study were restricted to asking learners to specify their L2 problems while composing. Brief, simple instructions requesting learners to report specific problems rather than general language areas worked well and facilitated data analysis.

Noticing sheet 2: NS-2 was designed for the comparison stage when reformulated or corrected essays were compared with learners' original writing. This sheet encouraged input noticing and processing of feedback. The instructions asked the learners to select at least five language features they considered important, and I added the question: "Is there anything in your reformulated text that attracts your attention? Anything that stands out as interesting or strange? If so, what?", adapted from (Park, 2011). Asking learners to select five things that attracted their attention was the way I settled on to cope with comprehensive feedback and still allow for learner-initiated noticing, something especially difficult in open 300-word writing tasks. Not asking learners to select features would have made their feedback analysis unduly time-demanding, attention span might have gone down and reports could also have been difficult to analyse. This issue is further discussed in Chapter 9 (strategies to cope with noticing and learner-initiated noticing).

9.3.3 Exit questionnaire

A supplementary exit questionnaire was used to collect information about learners' (Q1) impressions of the advantages and disadvantages of the ECC they experienced; (Q2) opinions on the usefulness of noticing sheet-1; (Q3) criteria to select what to focus on in their feedback; (Q4) suggestions to make error correction efficient. The questionnaire elicited quantitative (via multiple-choice questions) and qualitative (via 'why?' open-ended questions) data.

The decision to include an exit questionnaire in the design might be criticised. It has the disadvantages of offline data collection procedures and there was a one-week lapse between the treatment and questionnaire delivery. Leow (2013) points out that "Off-line measures at the stage of post-test can only make inferences as to whether the learner paid attention to or noticed targeted forms in the input during the experimental exposure" (pp. 13–14). In this sense, the exit questionnaire is a coarse-grained measure of attentional processes that cannot really determine if the results obtained are as a consequence of noticing. Including the questionnaire in a different stage would, however, have influenced the participants' performance in subsequent stages. There were only two options: placing it at the end of the design or abandoning it. I chose the first option. The usefulness of post-exposure questionnaires depends on learners' capacity to separate what they notice during learner-stimuli interaction from what they notice while completing the questionnaire (Mackey, 2006; Robinson, 1995). However, asking learners what they think about what was happening during experimentation is a way of enquiring into learners' own perceptions. Murphy (1993) stresses that learners' perceptions must be valued.

The aims, research questions and data collection methods are presented together in Table 13, below.

Table 14 – *Aims, research questions and data collection methods.*

Aim of the study	Research question	Data collection method(s)
Aim 1: Explore the effects (if any) of different comprehensive (eventually replaced by semi-comprehensive) ECCs (DIR, REF, SELF) on learner-initiated noticing at two stages (composing and feedback) of the writing process.	**RQ-1:** What are the effects (if any) of different comprehensive (eventually replaced by semi-comprehensive) ECCs (DIR, REF and SELF) on learner-initiated noticing at two stages (composing and feedback) of the writing process?	**NS-1 AND NS-2** − Occurrence of noticing − Amount of noticing − Error types noticed
- Whether different comprehensive (eventually replaced by semi-comprehensive) ECCs (DIR, REF, SELF) lead to leaner-initiated noticing in the composing and feedback stages, i.e. The occurrence of noticing reports.	**1.a -** Do different comprehensive (eventually replaced by semi-comprehensive) ECCs (DIR, REF and SELF) lead to learner-initiated noticing in the composing and feedback stages? i.e. The occurrence of noticing reports.	
- If so, the study would aim to examine the amount of noticing and the error types (grammatical or non-grammatical) learners pay attention to by themselves in each ECC at the composing and feedback stages, i.e. The number of noticing reports and error types reported as noticed.	**1.b -** If so, how much noticing and what error types (grammatical, non-grammatical) do learners pay attention to by themselves in each ECC at the composing and feedback stages, i.e. The number of noticing reports and error types reported as noticed.	
Aim 2: Explore the effects (if any) of the above noticing-supported ECCs at the feedback stage on learners' writing accuracy in the rewriting and new writing stages.	**RQ-2:** What are the effects (if any) of the above noticing-supported ECCs at the feedback stage on learners' writing accuracy in the rewriting and new writing stages?	**ESSAY ACCURACY ACROSS TESTS** **Occurrence of effects (**of noticing-supported ECCs in the feedback stage**)** on learners' writing accuracy in the rewriting and new writing stages. **Error types** (grammatical or non-grammatical) that were more amenable to correction in different noticing-supported ECCs.
- Whether the input provided by different (the ones above) noticing-supported ECCs at the feedback stage has any effects on learners' writing accuracy in the rewriting and new writing stages, i.e. The occurrence of effects of noticing-supported ECCs in the rewriting and new writing stages.	**2.a -** Does the input provided by the above noticing-supported ECCs at the feedback stage have any effects on learners' writing accuracy in the rewriting and new writing stages? i.e. The occurrence of effects of noticing-supported ECCs in the rewriting and new writing stages.	
- If so, the study would aim to examine what error types (grammatical or non-grammatical) are more amenable to correction in different noticing-supported ECCs, i.e. The error types learners corrected by themselves in different ECCs.	**2.b -** If so, what error types (grammatical or non-grammatical) are more amenable to correction in different noticing-supported ECCs? i.e. The error types learners corrected by themselves in different ECCs.	

Note. ECCs = Error Correction Condition(s), *DIR* = Direct group, *REF* = Reformulation group, *SELF* = Self-correction group, *NS* = Noticing Sheet

9.4 The design

Seeking answers to the aforementioned RQs, a four-stage writing task (composing/ error correction-noticing/ rewriting/ new writing) was designed. Learner-initiated noticing processes (occurrence and effects) constituted a common thread during the writing task.

Stage 1 (composing or pre-test): Stage 1 acknowledged the role of output in SLA (Swain, 1985, 1993, 1995) and explored learner-initiated noticing during L2 production via NS-1 (Appendix F). For writing task 1 see Appendix H.

Stage 2 (error correction-noticing): Stage 2 acknowledged the role of input (Krashen, 1985) and feedback in SLA and explored learner-initiated noticing in WCF. The 'error correction-noticing' stage engaged learners in analysis of their feedback by comparing their original essay with its reformulated or error corrected version. Learner-initiated noticing was prompted via NS-2 (Appendix G). Learners in the self-correction condition were engaged in analysing their own essays. Noticing opportunities added to the input provided by different ECCs constituted the treatment.

The effects of noticing-supported ECCs on learners' writing accuracy were analysed at two points:

Stage 3 (rewriting or post-test): Rewriting or writing task 2 (Appendix I) allowed learners to engage with feedback and test the effects of noticing-supported ECCs immediately after they were provided.

Stage 4 (new writing or delayed post-test): The new task or writing task 3 (Appendix J) tested the effects of noticing-supported ECCs beyond rewriting.

9.5 Treatments

Experimental group 1 (DIR): NS-1, NS-2 and direct error correction. Learners completed NS-1 (composing stage), NS-2 (feedback stage) and received direct error correction. The researcher corrected all types of errors in the essays and provided the correct form.

Example:

Learner's text

Physical appearance for me its important cause our physical appearance its a reflection of what we are. If you excersice your body segregates hormones that makes you fell happy so, worry about your physical image could have good consequences.

Direct error correction:

Physical appearance ~~for me~~ ~~its~~ is important <u>for me</u> ~~cause~~ <u>because</u> our physical appearance ~~its~~ is a reflection of what we are. If you ~~excersice~~ <u>exercise</u>, your body produces hormones that ~~makes~~ <u>make</u> you ~~fell~~ <u>feel</u> happy, so ~~worry~~ <u>worrying</u> about your physical image could have good consequences.

Experimental group 2 (REF): NS-1, NS-2 and reformulation error correction. Learners completed NS-1 (composing stage) and NS-2 (feedback stage) and received their reformulated essay. Three native speakers reformulated the learners' essays.

Example:

Learner's text

It is known that some the biggest enterprices in the world spend more money in their publicity campaigns rather than the cost of their products. Their marketing department is composed by people whom knows the psycological behaviour of most of the people.

Reformulation:

It is well known that companies spend far more on their advertising campaigns than they do on the actual products they sell. Marketing departments are typically made up of people who understand human psychology and how it affects people's behaviour.

Experimental group 3 (SELF): NS-1, NS -2 and self-correction. Learners completed NS-1 (composing stage) and self-corrected using NS-2. This group allowed me to differentiate the effects of error correction conditions from the effects of noticing opportunities.

Example:

Learner's text:

The media have benn an instrument to place people in an irreal world where every one is obliged to act as a beauty model.

Learner's self-correction:

The media have been an instrument to place people in an unreal world where everyone is obliged to act as a beauty model.

Control group (CONTROL): No NSs, no WCF. Learners received no error correction and were not offered any noticing opportunity; nonetheless, they wrote writing task 1, rewrote writing task 1, now called WT-2, and wrote a new essay (WT-3).

9.6 Experimental procedure

All tasks took place during class periods and were set up by the researcher. The researcher corrected essays in DIR-G; three previously trained native speakers (the same in pilot study) reformulated the essays in REF-G. To avoid onerous pen-and-paper work, experimental sessions were performed in a computer room especially adapted for this study. Learners typed their essays on computers (grammar and spell-check functions deactivated and no Internet access). Each student had their own folder in their personal computer. After each session, students' work was sent to the researcher's computer via a server. The design required three two-hour sessions: session one (stage 1), session 2 (stages 2 and 3), session 3 (stage 4). Table 15 was constructed to facilitate the reader's comprehension of the experimental procedure. The reader is advised to refer to Table 15 as they read the procedure's description.

Table 15 – *Experimental procedure.*

Group	Week 1		Week 3			Week 4	
	SESSION 1 (Stage 1)		**SESSION 2 (Stages 2 and 3)**			**SESSION 3 (Stage 4)**	
	First hour	**Second hour**	**First hour**		**Second hour**	**First hour**	**Second hour**
	Composing Pre-test *Essay 1*	*Noticing sheet 1*	Error condition	*Noticing Sheet 2* Students choose at least five language features	**Rewriting** Post-test *Essay 2*	**New-writing task** Delayed post-test *Essay 3*	Exit questionnaire
			Treatment				
DIR *N* = 15	✓	✓	✓ Direct error correction	✓	✓	✓	✓
REF *N* = 15	✓	✓	✓ Reformulation	✓	✓	✓	✓
SELF *N* = 15	✓	✓	✓ Self-correction	✓	✓	✓	✓
CONTROL *N* = 15	✓	x	No error correction	x	✓	✓	x
60 participants	60	45		45	60	60	45

Note. DIR = *DIRECT* group, *REF* = Reformulation group, *SELF* = Self-correction group

Session 1: stage 1 – composing or pre-test (week 1): During this composing stage, the four participant groups completed writing task 1, an opinion essay on a given prompt. CONTROL-G experienced no treatment (no error correction or noticing sheets). Instead, the learners wrote essay 1 (session 1), re-wrote essay 1 (session 2) and wrote a new essay (session 3). CONTROL-G received no WCF other than the researcher's general comments on the content of their writing. Five short story books were raffled among the participants in CONTROL-G to thank them for their participation.

The DIR, REF and SELF groups received a writing prompt with instructions for writing task 1 and noticing sheet-1. Learners were informed about task requirements: length, time available and no dictionary support. Students had a two-hour session to write and proofread their essay before submitting it. The DIR, REF and SELF groups were invited to write down their L2 problems on noticing sheet-1. This sheet was the same for all three experimental groups and was completed during the same two-hour session.

At the end of the session, essays (in the four participant groups) and noticing sheet-1 (in the three experimental groups) were sent to the researcher's folder. Each essay was printed twice (once for error analysis by the researcher, and once for the students at the noticing stage). Fifteen essays from the DIR-G were corrected by the researcher. Fifteen essays from REF-G were divided into three sets. Each native speaker received a set of essays and had ten working days to reformulate them. Corrections in the DIR and REF groups were made electronically by the researcher and by the native speakers, respectively. Only printed versions were returned to students for their error inspection in the noticing stage.

Session 2 (first hour): stage 2 – error correction – noticing or treatment (week 3): The treatment session (session 2) was held 12 days after session 1 to allow time for essay correction and reformulation. During this stage:

Learners in the DIR and REF groups received their printed original essay and were given time to read it and recall what they had written. DIR-G received their error corrected essay and noticing sheet-2 (Appendix G), whereas REF-G received their typed reformulated essay and the same noticing sheet-2. The DIR and REF groups compared their feedback with their original writing and completed noticing sheet-2. This uncontrolled condition aimed to collect evidence about what learners noticed (if they did so) by themselves, as noticing sheet-2 was an almost blank sheet of paper with no teacher's guidelines on what to look at.

Learners in SELF-G received their printed original essay with no alterations and the same noticing sheet-2 (Appendix G) that the DIR and REF groups received. Learners reread their essay, identified their own mistakes and corrected them. Detected mistakes and corrections were reported on noticing sheet-2.

Learners in CONTROL-G neither received their printed original essay nor engaged in feedback analysis. Instead, learners rewrote a second essay with the same prompt as in essay 1. Consequently, this group finished one hour before the others. All experimental groups had one hour to analyse their feedback and complete noticing sheet-2. CONTROL-G moved directly to rewriting essay 1; learners in all groups had the same amount of time (one hour) for rewriting. Time on task was tested in the pilot study where there was no time limit, one hour proved sufficient for the task.

Session 2 (second hour): stage 3 – rewriting or post-test (week 3): Stage 3 took place in the second hour of session 2. CONTROL-G did not participate in the

second hour of session 2 since they had no treatment. This group worked on stage 3 during the first hour of session 2.

After one hour of noticing in the three experimental groups, all materials (original essay, noticing sheet-2, corrected/ reformulated essays) were collected. The DIR, REF and SELF- groups received writing task 2 (same writing prompt as for essay 1, now called essay 2). Students rewrote their essay (see Appendix I) on their computers with no support (reformulated/ corrected essay, noticing sheets or dictionaries). Students did not know about this post-test or the new writing post-test. The word rewriting is used instead of revising as 'revising' may imply having access to received error correction, which was not the case (all corrected and self-corrected essays were previously collected). Rewriting the essay was included as a way to engage learners with the feedback received and test whether there was any immediate improvement in accuracy. At the end of session 2, rewritten essays in the experimental and control groups were sent to the researcher's folder.

Session 3: Stage 4 – new writing or delayed post-test (week 4): One week after the rewriting session, all groups (three experimental and one control) wrote a new writing task (see Appendix J) on a new but similar topic (same topic for all groups). One hour was allowed to do this. During the same session all groups completed an exit questionnaire (see Appendix K).

9.7 Methodological strategies

This section aims to bring together the experimental manipulations that I prefer to call methodological strategies. They were introduced in previous sections: 4.3.1 and 9.3.2. These strategies were utilised to deal with complicated issues, such as comprehensive EC, noticing and learner-initiated noticing. Bringing them all together

before moving on to the data analysis may assist comprehension and add clarity to my design.

9.7.1 Strategies to cope with comprehensive EC

Originally I aimed to examine comprehensive or unfocused EC. Not only are studies in comprehensive EC necessary, they are also more classroom representative. In Chapter 7, however, I explained the methodological challenges this EC technique would entail: dealing with teacher-raters' beliefs and everyday practices concerning what an error is, what is or is not acceptable, deciding on the linguistic category each feature belonged to. The first and most important methodological manipulation in my design was moving from comprehensive to semi-comprehensive EC. Learners did receive comprehensive EC during their feedback and noticing opportunities. This was crucial for learner-initiated noticing, which implied all error types (not exclusively predetermined linguistic targets) were corrected. It was the essays' error analysis that was semi-comprehensive, i.e. accuracy was measured by considering only seven selected linguistic features: spelling and lexis (non-grammatical accuracy), omission of constituent, unnecessary definite article, S+V agreement, 3rd person singular and gerunds (grammatical accuracy).

The need for inter- and intra-reliability in accuracy analyses led me to shift to semi-comprehensive error analysis. An examination of six studies (Ellis et al., 2008; Ferris, 2006; Sheen et al., 2009; Truscott & Hsu, 2008; Van Beuningen et al., 2008, 2012) claiming to work with comprehensive EC in the literature review was actually done after implementation of the design. I searched for explanations to problems I was unable to foresee before undertaking comprehensive EC. This methodological manipulation should not be considered a failure but rather an opportunity to look deeper into comprehensive error correction studies. As explained above, learner-

initiated noticing in WCF required not only providing learners with freedom to choose what they wanted to notice but also with correction of all error types so they could actually choose what they wanted to pay attention to. In Chapter 7, I reported the challenges, strategies, strengths and limitations of the six studies I found in WCF dealing with comprehensive EC. Different criteria for error categorization and different operationalisations of 'unfocused' WCF (usually referring to different ranges of errors) characterize all of them. Few details on the rubrics or guidelines employed to categorise error categories are provided. Van Beuningen et al.'s (2012) study is the only one that distinguishes grammatical from non-grammatical errors. Spelling, vocabulary, mechanics and punctuation are treated differently in different studies. For instance, some exclude lexis in the unfocused group (Ellis et al., 2008; Sheen et al., 2009); others exclude lexis and mechanics but include spelling (Truscott & Hsu, 2008); some others refer broadly to grammatical and ungrammatical errors without specifying how lexis and other features that might have multiple functions are dealt with (Van Beuningen et al., 2012). Targeted errors in all of these studies rely on instructors' inferences as to what they consider to be the most typical errors in the specific context of each one. The selected measured errors in the present study include grammatical and non-grammatical categories. The selected targeted errors resulted from both the raters' inferences on what they considered to be the most typical errors (as in most existing studies on comprehensive WCF) and the error rate analysis of essay 1. The error categorization criteria the raters and the researcher developed for this study are controversial and can be much improved, but they do constitute a methodological strategy. Unfocused studies in WCF are crucial to prove the efficacy of WCF. If a way to do research on unfocused WCF exists, researchers first have to agree on the operationalization of 'unfocused' and on how grammatical and non-

grammatical errors will be categorised. Comprehensive EC was not feasible with the open, uncontrolled, learner-centred, 300-word opinion essay I worked with. The combination of various uncontrolled conditions rather than the technique itself might have constrained the feasibility of comprehensive EC in this study.

9.7.2 Strategies to cope with noticing and learner-initiated noticing

(a) Noticing vs. reported noticing*:* The operationalization and measurement of 'noticing', 'learner-initiated noticing' and cognitive processes in general constitute one of the greatest challenges for SLA researchers. The use of noticing sheets to collect evidence of noticing has advantages and disadvantages. Advantages (already mentioned in Section 9.3.2) include: capturing attentional processes in real time, in a non-disturbing way and reasonably easily compared with other methods; allowing for learners' autonomous analysis of input; gathering information about where learners concentrate their attention and learners' nature of awareness. Concerning disadvantages, Mackey (2006) remarks that the use of [noticing] sheets (she calls them uptake sheets) to obtain introspective data on learners' noticing superficially connects noticing to the facts that prompted them. Leow (1997) and Robinson, Mackey, Gass and Schmidt (2011) also explain that input processing occurs momentarily, whereas diaries and uptake sheets last longer. Thus, there is the possibility of forgetting an experience at the time of reporting. This observation, added to Mackey's (2006) and Schmidt's (1990) remark that reported noticing does not always reflect everything people notice, makes it reasonable to use learners' "reported" noticing rather than learners' actual noticing to refer to learners' reports in NSs.

(b) Opportunity to leave NSs blank*:* 'Learner-initiated noticing' was defined as "episodes when learners pay attention, by themselves in the absence of any external

intervention, to new [or problematic] linguistic data in the input [and/or output], and relate these to existing knowledge". As already explained in Chapter 4, my aim to ensure that learners' noticing was self-initiated and not imposed validates the consent I gave them 'to leave the noticing sheets blank if they wanted to'. Learners' positive response to these specific instructions in the pilot study supported this decision.

(c) Five features to report*:* Asking learners to "select at least five things, more if they wanted to" from their feedback on NS-2 was a decision based on the pilot study experience. Not limiting learners' language features to those reported in their feedback during piloting led them to revise and report every change or correction made to their corrected or reformulated essays. Learners claimed everything was important for them. Unfortunately, participants' L2 proficiency and essay length led to reporting numerous errors which consequently reduced the time available for the rewriting stage. Limiting learners' reports aimed to help learners prioritize what attracted their attention the most and not making the task tiring. I chose five because during data analysis in the pilot study I realised that the top five to seven features in the error analysis and in the reported noticed features were the numbers I, as researcher, recalled without looking back at the results. I considered this number manageable for learners, raters and researcher and it suited the available time on-task.

Chapter 10: Data Analysis

One hundred and eighty 300-word essays produced by four groups in the composing (60), rewriting (60) and new writing (60) stages, 90 noticing sheets from all experimental groups noticing (sheet-1 (45) + noticing sheet-2 (45)) and data from 45 exit questionnaires constituted the data to analyse.

10.1 Essay analysis

Essay analysis required quantitative, descriptive and inferential analysis. As in the pilot study, learners' essays were first coded for linguistic errors. The analysis considered the first 300 words of each essay. If essays had more, only the first 300 were considered. Grammatical and non-grammatical error ratios ([number of linguistic errors/ total number of words] x 100) were computed for different dependent variables (i.e. overall, grammatical, non-grammatical accuracy) in the composing, rewriting and new writing stages. ANOVA and post hoc tests for statistically significant results were performed. Essay analyses for selected error types were also performed. Rubrics for error categorizations resulting from the pilot study were used.

10.2 Noticing sheets analysis

First, to collect evidence about learner-initiated noticing, learners were given freedom to select what to pay attention to. No pre-selected categories were established and learners were also free to decide whether to report or not. Data obtained from NSs went through qualitative analyses. I used Hanaoka's (2007) and Hanaoka and Izumi's (2012) categorizations: problematic features noticed (PFNs) for NS-1 and features noticed (FNs) for NS-2. First I reviewed data from the NSs to identify PFNs and FNs mentioned by at least one learner. The identified PFNs and FNS were then classified

as grammatical and non-grammatical error types. Within those types, further error categories were identified: spelling, vocabulary, punctuation and connectors (non-grammatical), omission of words, verb tenses, word order etc. (grammatical). The same error categorization created for essay analysis was used for noticing sheets. Error categories were modified (if relevant) based on the emerging data. For example, the category "content organization" was added. After all error categories had been established within grammatical and non-grammatical categories, the NSs were re-examined, and one point was added to the category each time an item was reported by a participant. See examples of the analyses performed below. Examples include participants in the three experimental groups: DIR, REF and SELF, and reports of PFNs on NS-1 (examples 1, 2 and 3) and of FNs on NS-2 (examples 4, 5, and 6).

Example 1

Information as reported on NS-1 – Student 13 (REF)

"I had problems to write my ideas in order. First, I wrote the second paragraph as an introduction but it did not make sense. Therefore, I wrote another paragraph as an introduction. Eventually, I didn´t know if my sentences in third paragraph made sense."

Analysis and counting

Learner's notes above refer to the same problem: text construction. Counting was reported as follows:

Non-grammatical: 1 → Content organization (1)

Example 2

Information as reported on NS-1 – Student 20 (DIR)

“I had problems with synonyms, for example I didn´t know the synonym for affective, aesthetic, appearance.”

Analysis and counting

Whenever learners reported specific examples for a particular error category, each example was counted as one. For example, in the report above, the learner signalled three examples of synonyms he had problems with. Thus, the synonym category scored three points. Synonyms in turn belong to the vocabulary error type. Counting was reported as follows:

Non-grammatical: 3 → Vocabulary (3): lack of synonyms

Example 3

Information as reported on NS-1 – Student 12 (SELF)

Problems with the spelling of: *healty, *anorexya, *nutricion.

Problems with tenses, for example: I am not sure this is correct “It has being changing.” I have also forgotten about several connectors.

Analysis and counting: Counting was reported as follows:

Non-grammatical: 4 → Spelling (3), Connectors (1)

Grammatical: 1→ Tenses (1)

Example 4

Information as reported on NS-2 – Student 4 (REF)

“I omitted words that are important to make my message clear.”

Analysis and counting: Counting was reported as follows:

Grammatical: 1 → Omission of words (1)

Example 5

Information as reported on NS-2 – Student 2 (DIR)

“Many things were corrected in my essay. The most important thing to me was how ideas were expressed; the order of the words in a sentence is different!”

Analysis and counting: Counting was reported as follows:

Grammatical: 1 → Word order (1)

Example 6

Information as reported on NS-2 – Student 15 (SELF)

Original: If you use the most expensive bag, you are the perfect person in the world. This is sad becose people worth about what they have not how are they.

Correction: If you wear the most expensive bag, you are the perfect person in the world. This is sad because people worry about what they own not about who are they.

Analysis and counting: Counting was reported as follows:

Non-grammatical: 2 → lexis (1): use vs. wear + spelling (1): becose vs. because

Grammatical: 1 → syntax (1): people worth about what they have not how are they vs. people worry about what they have not about how are they.

Table 15 shows examples of complete analyses of NS-1; Table 16 shows examples of complete analyses of NS-2. The total of PFNs and FNs in each NS was recorded, as was the number of errors reported within each error type (GR and N-GR) and error category (vocabulary, punctuation, tenses, prepositions, etc.). In the end, each NS had a sub-total of GR and N-GR errors.

Table 16 – *Examples of noticing sheet 1 analyses in experimental groups.*

Student	Information as reported on noticing sheet	Counting
Student 12 (DIR)	-I had problems with synonyms, for example I didn´t know the synonym for "affective, aesthetic, appearance". -I had problems with linkers. I translated the linkers from my native language into English, e.g. "Sin embargo, aunque, porque, a pesar de." -I overused "that". This was because of my lack of vocabulary. I do not know words to link sentences. -I had problems to write my ideas in order. First, I wrote the second paragraph as an introduction but it did not make sense. Therefore, I wrote another paragraph as an introduction. Eventually, I didn´t know if my sentences in third paragraph made sense. -I forgot to write commas after adverbs and linkers.	Non-grammatical features: 10 *Vocabulary (3)* Lack of synonyms for "affective, aesthetic, appearance". *Connectors (4)* 'Sin embargo, aunque, porque, a pesar de.' *Word repetition-overuse of 'that' (1)* *Content organization (1)* *Punctuation (1)* Grammatical features: 0
Student 13 (REF)	-I had problems to start the essay, I knew the idea I wanted to write, but I didn´t know how to start. -I miss synonyms for some words. -Problems to move from general to specific ideas keeping coherence in the text. -Problems to use formal language -I hesitated with irregular plurals -Tenses was another problem. I have to think about the tense I need past, future etc.	Non-grammatical features: 5 *Content organization or lack of writing techniques to:* -start texts, paragraphs (1) -organize main vs. secondary ideas (1) -move from one idea to another keeping coherence in the text (1) *Lack of synonyms (1)* *Lack of formal language (1)* Grammatical features: 2 Tenses (1) Irregular plurals (1)
Student 12 (SELF)	-Problems with the spelling of: *healty, *anorexia, *nutricion. I was unable to recall the spelling. -Problems with tenses, for example: I am not sure this is correct "It has being changing." -I have also forgotten about several connectors.	Non-grammatical features: 4 Problems to recall the spelling of some words (3): *healty, *anorexia, *nutricion I forgot about the meaning of different connectors (1) Grammatical features: 1 Hesitations with the use of tenses (1) e.g. "It has being changing." I'm not sure if it is correct.

Table 17 – *Examples of noticing sheet 2 analyses in experimental groups.*

Student	Information as reported on noticing sheet	Counting
Student 2 (DIR)	*Vocabulary:* good/well, true/real, see/realize, seem to be/look, conduces/leads, like/such as *Spelling:* *aperance/appearance, *atractiv/attractive, *dissese/disease *Connectors:* However *Capital letters:* *Beaty/beauty, Programs/programs Many things were corrected in my essay. The most important thing to me was how ideas were expressed; the *order of the words* in a sentence is different!	12 total non-grammatical features: Vocabulary (6) Spelling (3) Connectors (1) capitalization (2) 1 grammatical feature: Word order (1)
Student 4 (REF)	If I increase my *vocabulary*, I will be able to avoid *repeating the same words* in the same paragraph. I need to write *each idea in a separate sentence* as in the reformulated essay. In the original, I tried to say two things in the same sentence. Indeed, however and other *connectors* are important to link my ideas and sentences. I *omitted words* that are important to make my message clear. I have to think as an English speaker to structure my ideas better. At present, I *translate* what I write.	4 non-grammatical features: Vocabulary (1) Content organization (2): Repetition of ideas & one sentence per idea Connectors (1) 1 grammatical feature: Omission of words (1) *Note:* Thinking in English to avoid direct translation was not classified in any category. This was actually identified as common emerging findings in all experimental groups.
Student 15 (SELF)	*Original:* If you use the most expensive bag, you are the perfect person in the world. This is sad *becose people worth about what they have not how are they. *Correction:* If you *wear* the most expensive bag, you are the perfect person in the world. This is sad *because people worry about what they have not about how are they.* What I tried to say was 'Es triste que la gente no se preocupe por lo que es.' But I think I contradicted myself. This happened in the moment my PC froze, I panicked and I missed the idea of what I was trying to say.	2 non-grammatical features: Lexis (1) Spelling (1) 1 grammatical feature: Syntax (1) *Note:* Learner's self-correction of syntax was partially corrected however; this was not considered in the features noticed count. What counted as noticing was learners' efforts to correct or explain the error

10.3 Exit questionnaire analysis

The exit questionnaire required quantitative analysis for multiple choice questions and qualitative analysis for 'why?' open-ended exit questions. Analysis of the latter was not problematic as the small number of students in each group (*N= 15*) allowed me to report all the students' answers.

Answers to Q1 (impressions about ECC) and Q2 (usefulness of NS-1) were classified as positive and negative impressions (Q1) and was yes-no answers (Q2). Totals under each category were summed and rated. In Q1, most students were firm in their positive or negative opinion about their experienced EC. However, when a learner's answer alluded to positive and negative feelings, he/she was categorized as "balanced". Common reasons for positive or negative impressions (Q1) or yes-no answers (Q2) were grouped together.

E.g.

DIR: Four students (2, 7, 13 and 14) remarked on the importance of making mistakes and the value of error correction.

REF: Its richness in vocabulary, formal language and content organization as additional to grammar accuracy was emphasised by six students (1, 9, 10, 11, 12, 14).

SELF: Three students (1, 6, 14) were categorised as balanced. They all agreed self-correction was a good technique, however incomplete.

Answers in Q3 (noticing criteria) and Q4 (suggestions to make WCF effective) also required identification of common features, but not positive or negative as in Q1 and Q2. Common criteria and suggestions, the ones that best reflected learners' responses, were created. For example, a learner's response to noticing criteria (Q3) "Things that were really hard to structure" and "Things that I was unable to express"

resulted in the creation of a “Sentences, words or structures learners struggled to formulate” category. Similarly, learners’ responses of “Words I did not know” and “Phrases I had never heard or seen” resulted in the creation of a “New words or expressions” category. Categories were added and modified based on information emerging from the data. Once the categories were established, the responses were re-examined and one point was added to each response that fell into the corresponding category. Totals under each category were summed and rated.

Chapter 11: Results

11.1 Essay results

Written essays were used to measure the effects of different noticing-supported error correction conditions (DIR, REF, SELF and CONTROL) on performance accuracy (non-grammatical, grammatical, overall) across pre-, post- and delayed post-tests. Accuracy in essay writing was operationalized as the percentage of correct usage of grammatical and non-grammatical features. The number of words (300), genre (opinion essay), time (1 hour) for essay writing and marking procedures were controlled in the three tests.[13] Error categorization criteria (Section 8.2) for raters were provided and they participated in prior training too.

Essay 1 (overall accuracy) was used as a base to find out whether the four participant groups were comparable. When students wrote essay 1 they were in equal conditions, no group had experienced any treatment. The error rate in each group in essay 1 was calculated; see Table 18, below.

Table 18 – *Overall groups' performance in essay 1.*

Group	Mean number of errors	Standard deviation
DIR	30.1	10.0
REF	31.9	11.0
SELF	32.0	6.8
CONTROL	28.1	9.1

Note. DIR = Direct group, REF = Reformulation group, SELF = Self-correction group

Then, a one-way ANOVA on the test 1 results showed no significant differences between groups: $F(3, 56) = 0.59$, $MSE = 52.1$, $p = 0.63$. This means the groups were comparable. Thus, any improvement from the post-tests was not as a consequence of prior differences between groups.

[13] Test and essay refer to the same thing and are used interchangeably.

11.1.1 Descriptive statistics

The four groups' accuracy performance across the three essays is presented in Tables 18 and 19, below. Numbers in parentheses indicate the average number of errors for overall accuracy (Table 19), and for non-grammatical (N-GR) and grammatical (GR) accuracy (Table 20).

Table 19 – *Overall accuracy performance of the four groups across tests.* Numbers represent mean number of errors (with SDs in parentheses).

Group	Essay 1 (Pre-test)	Essay 2 (Post-test)	Essay 3 (Delayed post-test)
DIR	30.1 (10.0)	22.0 (7.2)	27.3 (5.8)
REF	31.9 (11.0)	27.4 (12.0)	31.9 (14.3)
SELF	32.0 (6.8)	30.1 (6.9)	28.1 (9.3)
CONTROL	28.1 (9.1)	27.7 (10.2)	26.4 (13.5)

Note. DIR = Direct group, *REF* = Reformulation group, *SELF* = Self-correction group, *SD* = Standard Deviation

As observed in Table 19, all groups showed overall accuracy improvement from essay 1 to essay 2 (the numbers of errors in the four groups went down: from 30.1 to 22 for DIR; from 31.9 to 27.4 for REF; from 32 to 30.1 for SELF; from 28.1 to 27.7 for CONTROL). However, from essay 2 to essay 3, the number of errors in DIR and REF went up again: from 22 to 27.3 for DIR; from 27.4 to 31.9 for REF. The number of errors in SELF (with self-provided or self-initiated feedback and noticing opportunities) and CONTROL (no feedback, no noticing opportunities) from test 2 to test 3 again went down: from 30.1 to 28.1 for SELF; from 27.7 to 26.4 for CONTROL.

Table 19 displays the immediate overall accuracy improvement observed in all participant groups, it was partially retained until test 3 in the two groups receiving external explicit feedback (DIR and REF groups). Mean error in the DIR and REF groups went up again in test 3. However, the mean error seen in test 3 was still lower than in test 1 (for DIR) or equal to test 1 (for REF). SELF-G (with self-provided or

self-initiated feedback and with noticing sheets), however, was the only experimental group that improved its accuracy across the three essays: 32.0 (essay 1), 30.1 (essay 2), 28.1 (essay 3). Concerning CONTROL-G, (no feedback, no noticing opportunities), its performance was similar to that of SELF-G, i.e. its overall accuracy improved across the three tests: 28.1 (essay 1), 27.7 (essay 2), 26.4 (essay 3).

Table 20 – *N-GR and GR accuracy performance of the four groups across the tests.* Numbers represent mean numbers of errors (with SDs in parentheses).

Group	Essay 1 (Pre-test)		Essay 2 (Post-test)		Essay 3 (Delayed post-test)	
	N-GR	GR	N-GR	GR	N-GR	GR
DIR	18.4 (5.7)	11.7 (6.2)	11.6 (5.2)	10.4 (5.5)	13.8 (6.7)	13.5 (5.0)
REF	18.0 (6.9)	13.9 (6.9)	17.5 (8.0)	9.9 (4.8)	16.3 (9.2)	15.6 (7.8)
SELF	19.8 (4.5)	12.2 (5.5)	17.7 (4.9)	12.5 (6.7)	15.7 (3.8)	12.3 (8.3)
CONTROL	15.7 (5.9)	12.4 (6.4)	16.2 (6.9)	11.5 (5.1)	12.6 (5.6)	13.8 (8.9)

Note. N-GR = Non-grammatical, GR = Grammatical, SD = Standard deviation, DIR = Direct group, REF = Reformulation group, SELF = Self-correction group.

The data in Table 20 show that non-grammatical features mostly determined the trend described above for the overall accuracy performance of the four groups. N-GR errors went down from essay 1 to essay 2 in the DIR (from 18.4 to 11.6) and REF (from 18.0 to 17.5) groups, though N-GR errors went up again in test 3 for DIR (from 11.6 to 13.8). The REF group was the exception as, different from the overall accuracy performance, the number of N-GR errors kept going down: from 17.5 (essay 2) to 16.3 (essay 3). N-GR errors in the SELF and CONTROL groups, as in overall accuracy performance, kept going down across the three tests.

Figure 1 illustrates the groups' overall accuracy performance across the tests. Note that the numbers of errors in the DIR and the REF groups go down from test 1 to test 2, though they go up again from test 2 to test 3. The numbers of errors in the SELF and the CONTROL groups, conversely, go down across the three tests.

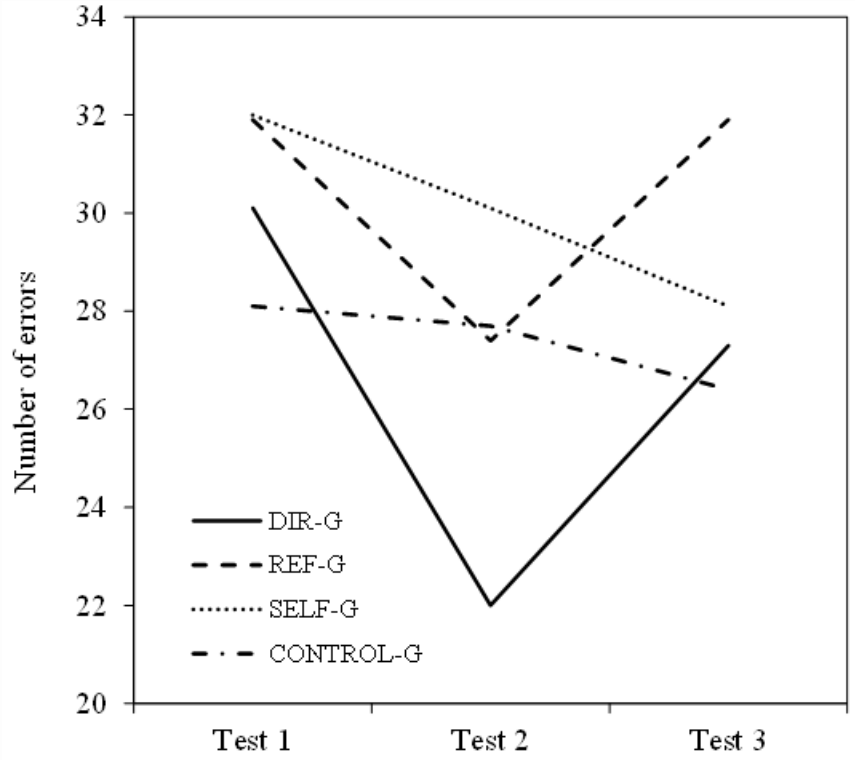

Figure 1 – Groups' overall accuracy performance across tests.

11.1.2 Inferential statistics

Inferential analyses were performed to determine if the results in Table 19 and Table 20 were statistically significant, i.e. if the results were not attributed to chance. Before that, I confirmed that my data met the assumptions[14] of an ANOVA. Figure 2 presents an example of normal distributions for the direct group. Similar distributions were found in the other three groups. Confidence intervals were calculated with $\alpha \leq 0.05$.

[14] (1) Groups must be independent of one another; the same data must not be contained in two groups; (2) the residuals (differences from the mean) must be approximately normally distributed; (3) the residuals must have approximately equal variances.

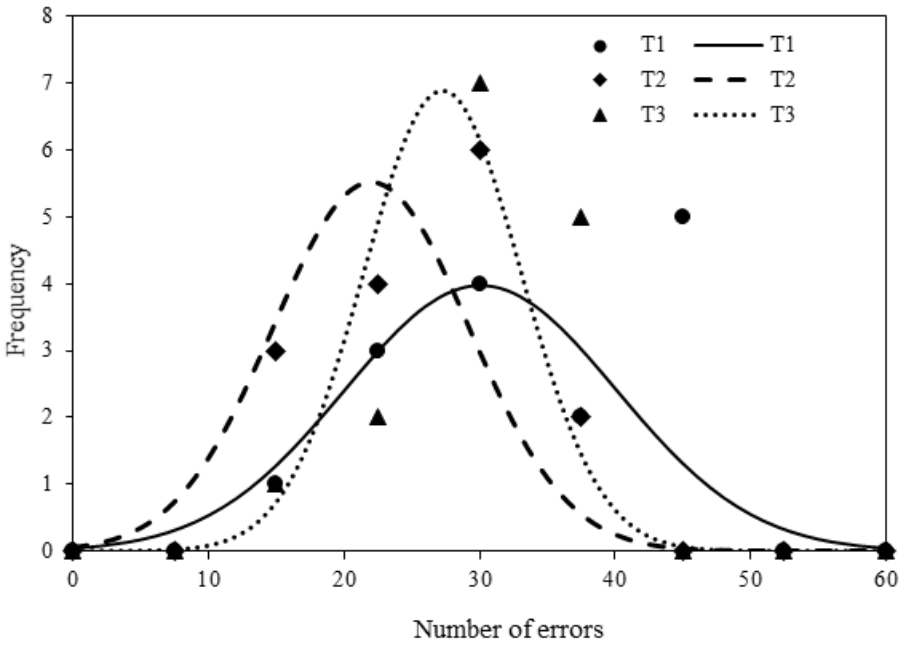

Figure 2 – Normal distributions for overall accuracy of DIR group across tests. Similar distributions were found for the other three groups.

Various one-way ANOVA tests were performed to measure the effects of different noticing-supported ECCs on learners' language accuracy within groups[15] (whether each group's accuracy performance improved, or not, across the three tests) and between them (comparing the four groups' accuracy performance in each test, i.e. whether groups continued to be comparable or not as they were in test 1).

The first questions I aimed to answer with each significance test were whether there were any differences in the overall, non-grammatical and grammatical accuracy performance of the DIR, REF, SELF and CONTROL groups across the three tests.

Significance tests within groups: Significant differences in overall accuracy performance from test 1 to test 3 only emerged for DIR-G. A one-way ANOVA showed the results in the three tests were statistically different for overall (*p = 0.03) and non-grammatical (*p = 0.01) accuracy, the latter influenced the former.

[15] The word 'groups' in 'within and between groups' does not refer to the four participant groups in the study. 'Groups' in ANOVA terminology refers to the performance of each of the four participant groups: (a) Across the three tests when it is read 'within groups', i.e. whether they improve or not; (b) In each of the tests when it is read 'between groups', i.e. whether they are comparable or not, as in test 1.

Significance was also found for non-grammatical accuracy in SELF-G (*p = 0.05), see results in Table 21, below. The grammatical category does not appear in this table, as it did not show significant results, i.e. it did not improve.

Table 21 – *Participant groups showing statistical significance across tests (within groups).*

Group	Overall				Non-grammatical			
	F(2, 42)	*MSE*	*p*	*Cohen's d*	*F*(2, 42)	*MSE*	*p*	*Cohen's d*
DIR	4.05	251.6	*0.03	0.11	5.25	180.6	*0.01	0.12
REF	0.65	102.8	0.53	- - -	0.17	11.1	0.86	- - -
SELF	0.96	58.1	0.40	- - -	3.16	62.1	*0.05	0.10
CONTROL	0.09	11.4	0.92	- - -	1.50	56.6	0.25	- - -

Note. MSE = mean-square error, *DIR* = Direct group, *REF* = Reformulation group, *SELF* = Self-correction group.
Statistical significance means improvement; here, *p indicates significance was found at $p \leq 0.05$
Grammatical category not included in this table as there were no statistical differences.

Post hoc *Cohen's d* tests showed that the significance values reported are at a 0.1 level of significance. In Cohen's terminology, this means a small effect size.[16]

Regarding the REF, SELF and CONTROL groups, ANOVA analyses within them revealed their overall accuracy performance across the three tests was statistically the same: REF = (p= 0.53), SELF = *(*p= 0.40), CONTROL = *(*p= 0.92). This means the REF, SELF and CONTROL error correction conditions had no significant effects on (or led to no improvement in) learners' overall accuracy performance across the three tests. Visible differences in descriptive statistics were not statistically significant.

T-tests on significant results: Next, in order to find out whether the significant test differences found in DIR-G (overall and non-grammatical accuracy)

[16] Cohen suggests that *d = 0.2* is considered a 'small' effect size, 0.5 represents a 'medium' effect size and 0.8 a 'large' effect size. A small effect size is one in which there is a real effect, i.e. something is really happening in the world which you can only see through careful study.

and SELF-G (non-grammatical accuracy) lay (a) among all of the tests and between each other or (b) only in one of them, further significance tests were run.

For the DIR group's overall accuracy results, T-tests[17] showed test differences lay in test 2. The results of tests 1 and 3 were the same and these equal results were different from the results of test 2. The results of test 2 were better (lowest error mean 22) than the results of tests 1 (mean error 30) and 3 (mean error 27). Details of test comparisons are in Table 22, below.

Table 22 – *Test comparisons in direct group (overall accuracy).*

Test comparison (error mean)	*F*(1, 28)	*MSE*	*p*
(30) T1 – **T2** (22)	6.39	488.0	*0.02
(30) T1 – T3 (27)	0.88	58.8	0.36
(22) **T2** – T3 (27)	4.86	208.0	*0.04

Note. MSE = mean-square error
**p indicates significance was found at $p \leq 0.05$*

For the DIR-G non-grammatical accuracy results, T-tests showed test differences lay in test 1. The results of tests 2 and 3 were the same and these equal results were different from the results of test 1. The results of test 1 were worse (highest error mean 18) than the results of tests 2 (mean error 12) and 3 (mean error 14). Details of test comparisons are in Table 23, below.

Table 23 – *Test comparisons in direct group (non-grammatical accuracy).*

Test comparisons (error mean)	*F*(1, 28)	*MSE*	*p*
(18) **T1** – T2 (12)	11.77	346.8	*0.03
(18) **T1** – T3 (14)	4.15	158.7	*****0.05
(12) T2 – T3 (14)	1.02	36.3	0.32

Note. MSE = mean-square error
**p* indicates significance was found at $p \leq 0.05$

For the SELF-G non-grammatical accuracy results, T-tests showed that despite a slight performance improvement being observed from test 1 to test 2 (mean error

[17] A t-test is considered to be a special case of one-way ANOVA. Whereas a t-test is limited to comparing the means of two groups, one-way ANOVA can compare more than two groups.

decreased from 20 to 18) and from test 2 to test 3 (mean error decreased from 18 to 16), these differences were not statistically significant. However, a statistical difference emerged between tests 1 and 3 (mean error decreased from 20 to 16). See details in Table 24, below.

Table 24 – *Tests Comparisons in SELF- Group (non-grammatical accuracy).*

Tests comparisons (error mean)	*F*(1,28)	*MSE*	*p*
(20) T1 – T2 (18)	1.53	34.1	0.23
T1 (20) – **T3** (16)	7.13	124.0	*0.01
T2 (18) – T3 (16)	1.46	28.0	0.24

Note. *MSE* = mean-square error
**p* indicates significance was found at $p \leq 0.05$

ANOVA tests showed there were no significant differences in the grammatical accuracy performance of any of the participant groups. Results for all groups in the three tests were statistically the same. Thus, no further significance tests were necessary.

Significance tests for selected error types: ANOVA tests were also run on the seven selected error types across the three tests. Table 25 shows the accuracy performance of the four participant groups for different error types improved exclusively for spelling in the DIR and CONTROL groups. Spelling was the only error type that showed statistical significance within groups, i.e. the only error type with improved accuracy across the three tests. This means error type did not influence accuracy performance (except for spelling in the DIR and CONTROL groups).

Table 25 – *Error types showing statistical significance within groups.*

	Spelling		
Group	*F*(2, 42)	*MSE*	*p*
DIR	3.48	76.8	*0.05
REF	0.02	0.6	0.99
SELF	0.80	13.1	0.47
CONTROL	3.86	59.5	*0.04

Note. MSE = mean-square error, *DIR* = Direct group, *REF* = Reformulation group, *SELF* = Self-correction group
**p* indicates significance was found at $p \leq 0.05$

Regarding statistical significance between groups, Table 26 shows that the error types (of the seven selected for this study) that showed statistical significance were: non-grammatical accuracy of T2 for the DIR group (*$p = 0.04$); spelling accuracy of T3 for the CONTROL group (*$p = 0.02$); lexis accuracy of T2 for the DIR group (*$p = 0.04$).

Table 26 – *Error types showing statistical significance between groups.*

Test	Non-grammar			Spelling			Lexis		
	F(3,56)	*MSE*	*p*	*F*(3,56)	*MSE*	*p*	*F*(3,56)	*MSE*	*p*
T1	1.31	44.2	0.29	1.07	21.2	0.38	1.49	16.2	0.24
T2 (DIR)	2.98	121.4	*0.04	1.79	38.9	0.17	2.98	47.5	*0.04
T3 (CONTROL)	1.02	44.6	0.40	3.91	83.0	*0.02	0.95	17.2	0.43

Note. MSE = mean-square error, *T* = Test, *DIR* = Direct group
**p* indicates significance was found at $p \leq 0.0$.
Only error types with significant results are included.

The question I wanted to answer with between groups analyses was whether there were any differences in the overall, N-GR, GR and error type accuracy performance of the DIR, REF, SELF and CONTROL groups in each of the three tests (i.e. whether the participant groups continue to be comparable/ equal as they were in test 1 or whether treatment has resulted in a change or made the groups different; if so, which group(s) and in which test(s)?). I already knew, because I compared the four groups before treatment, that, in Test 1, the four participant groups were comparable because no statistical differences were found among them before treatment. What I

discovered with the between tests analyses for T2 and T3 was that the participant groups were not comparable anymore in Test 2 and Test 3 because treatment did result in statistically significant differences.

The results in Table 26 show that the DIR group performed better than the other participant groups in test 2, especially in lexis, with its best performance in non-grammatical accuracy. These results confirmed the significant differences already found for the overall performance of the DIR group. They also show that the CONTROL group performed better than the other participant groups in spelling in test 3.

11.2 Noticing sheets results

Data collected on NSs allowed me to determine (a) the occurrence (or not) of both types of noticing; (b) the amount of reported information on each noticing sheet, and (c) the error types reported at the composing and feedback stages. Results are presented under each of these objectives.

11.2.1 Occurrence of both types of noticing

Noticing the hole (NS-1) and noticing the gap (NS-2) both occurred in the three experimental groups in the composing and feedback stages. Examples of phrases introducing evidence of learner-initiated noticing in this study were various: ‘A problem I had was’, ‘I cannot remember’, ‘It’s difficult for me to’, ‘I don’t know how to’, ‘My greatest challenge is’, ‘I forgot about’, ‘I hesitated with’, and ‘… is/are difficult for me’ on NS-1; ‘I learned’, I did not know’, ‘What attracted my attention was’, ‘I realized’ and ‘I would have never thought’ on NS-2.

On NS-2, students’ writing of their original and their reformulated, corrected or self-corrected sentence was already considered evidence of noticing. By doing this, learners’ demonstrated their ability to identify problematic features noticed and errors

or differences between essays 1 and 2. *Absence of noticing* (operationalized in this study as blank noticing sheets or no explicit report of awareness) also occurred on both noticing sheets. On NS-1, five learners (one in REF and four in SELF) handed in blank NSs. On NS-2, five learners in DIR-G handed in blank noticing sheets. Forty participants out of 45 in the composing stage completed NS-1. The same proportion completed NS-2 in the feedback stage. Table 27 summarises the above results.

Table 27 – *Occurrence and absence of noticing.*

Occurrence	Noticing	Absence
40 out of 45 participants	NS-1 Noticing the hole Composing stage	NS-1: 5 blank – REF (1) + SELF (4)
40 out of 45 participants	NS-2 Noticing the gap Feedback stage	NS-2: 5 blank – DIR (5)

Note. NS = Noticing sheet, REF = Reformulation group, *SELF* = Self-correction group, *DIR* = Direct group.

11.2.2 Amount of reported information on each NS

The results show that the three experimental groups differed in their amount of reported noticing in the composing and feedback stages. Table *28* presents a summary of the type and number of features reported on NS-1 and NS-2. As it is observed, reported noticing on NS-2 was more than reported noticing on NS-1.

Table 28 – *PFN on NS-1 and FN on NS-2.*

	NS-1 (PFN)			NS-2 (FN)		
Group	N-Gr	Gr	Overall	Non-Gr	Gr	Overall
DIR	47	5	52	49	34	83
REF	34	17	51	67	32	99
SELF	34	4	38	48	81	129
Total	115	26	141	164	157	311

Note. PFNs = Problematic features noticed; FNs = Features Noticed; NS = Noticing Sheet, DIR = Direct group, REF = Reformulation group, SELF = Self-correction group. N-Gr = Non-grammatical, Gr = Grammatical. There were four blank NS-1s in the SELF- group.

Table *28* indicates that the number of overall reported FNs on NS-2 (83 DIR, 99 REF, 129 SELF groups) was far larger than on NS-1 (52 DIR, 51 REF, 38 SELF groups).

Concerning NS-1, the results show that the numbers of overall problematic features noticed (PFNs) were almost the same in the DIR and the REF groups (52 vs. 51, respectively). The differences between these two groups lay in the higher number of non-grammatical features in DIR-G over REF-G (47 vs. 34, respectively). SELF-G, on the other hand, reported the least number of features in overall and grammatical accuracy; its non-grammatical accuracy was equal to REF group's. The four blank NS-1s in SELF-G might have determined this outcome.

The number of overall FNs reported in NS-2, contrariwise, showed that reported noticing in SELF-G was the highest. SELF-G reported 129 overall features vs. 99 reported in REF-G and 83 in DIR-G. The prevalence of grammatical features (double the number of non-grammatical features in the REF and DIR groups) is also noticeable.

11.2.3 Noticed error types

Noticing sheet-1: Concerning error types on NS-1, non-grammatical features were reported as the most problematic ones for learners in the three experimental groups: 115 non-grammatical vs. 26 grammatical ones (Table *28*).

Table 29 shows that, among non-grammatical features, lexis or lack of vocabulary emerged as learners' main problem. Spelling came up as the second problem closely followed by lack of connectors. Word repetition was fourth on the list. Learners explained that word repetition was as a consequence of their lack of vocabulary that obliged them to repeat the same words. Content organization was almost at the same level as word repetition. Unknown punctuation rules and limited

formal language were also reported in the experimental groups, but to a lesser extent. Notice that word repetition and formal language could both be classified within the lexis category. If so, lack of lexis would still be the top PFN in DIR-G and REF-G, the latter slightly higher than the former.

Learners stated on NS-1 that non-grammatical problems such as failure to start paragraphs or link ideas (Student 5-REF), lack of vocabulary (Student 15-REF), lack of synonyms (Student 6-REF; Student 13-REF) and spelling uncertainties (Student 15-REF) had led them to: (a) omit or modify what they were trying to say; (b) write incomplete sentences; (c) constantly repeat words. The little attention writing receives in their EFL classes, the necessity to get further practice in writing (Student 1-REF) and the challenge that thinking and writing in English entails (Student 7-DIR; Student 6-SELF) were also mentioned on NS-1. Finally, lexis and spelling were the only two features of the seven analysed in this study that learners considered problematic, 3rd person singular (3PS) was only mentioned by one learner in SELF-G.

Grammatical features were reported much less in the DIR and SELF groups (five and four, respectively) on NS-1 (Table *28*). REF-G showed higher reporting of grammatical features, mainly because of allusions to various tenses, prepositions and irregular verbs by some students.

Table 29 – *Type and number of PFNs reported as noticed on NS-1.*

Error type	DIR	REF	SELF
Non-grammatical			
Lexis	25	15	21
Spelling	9	5	7
Connectors	9	3	5
Word repetition	1	5	0
Content organization	1	3	1
Punctuation	1	2	0
Formal language	0	1	0
Contractions	1	0	0
Total non-grammatical	47	34	34
Grammatical			
Tenses	1	5	2
Prepositions	2	4	1
Irregular Verbs	0	4	0
Passive Voice	1	1	0
Others: Unnecessary words, 3PS, Possessive's, Irreg. plurals, conditional	1	3	1
Total grammatical	5	17	4
Overall	**52**	**51**	**38**

Note. *PFN* = problematic features noticed, *FN* = features noticed, *NS* = *noticing sheet*, *DIR* = direct group, *REF* = reformulation group, *SELF* = self-correction group, *3PS* = third person singular

Noticing sheet-2: Moving onto error types on NS-2, differences in the number of overall FNs between the DIR and REF groups were closer to each other (DIR 83, REF 99) when compared to the SELF group (129). SELF-G outnumbered the average for reported FNs where grammatical errors prevailed (see Table 30). SELF-G reported nearly twice the number of grammatical (81) over non-grammatical (48) errors when compared to the DIR and REF groups.

The opposite of SELF-G, the number of non-grammatical features reported was higher in the DIR and REF groups. Table 30 displays all selected errors for this study (lexis, spelling, omission of words, UDA, S+V agreement, gerunds and 3PS) that were reported on learners' NS-2. Non-grammatical errors such as lexis and spelling were frequently reported in DIR and REF groups, whereas grammatical

features such as omission of words were more frequently reported in SELF-G. Table 30 also indicates that text construction features, such as connectors, punctuation, content organization and formal language, were reported twice as often in REF than in other groups. The reformulation condition actually reported the highest amount of non-grammatical features. Non-grammatical features might have attracted learners' attention in the REF group because such features were completely new to them. New input might make it more salient. Learners in the REF (Student-8) and DIR groups (Student-1, Student-3) actually signalled new input had attracted their attention when comparing their original and reformulated/ corrected essays. Observe the contrast with the zero reporting of formal language in SELF-G.

Table 30 – *Type and number of features noticed on noticing sheet-2.*

This table shows that text construction and N-GR features* prevailed in the REF group whereas GR features** prevailed in the SELF group.

Error type	**DIR**	**REF**	**SELF**
Non-grammatical			
Lexis	25	32	18
Spelling	10	4	20
Connectors	5	11*	3
Punctuation	2	7 *	1
Content organization	1	5 *	1
Formal language	1	4 *	0
Unclear ideas, unplanned writing	3	1	1
Capitalization	2	0	1
Repetition of ideas	0	3	0
Contractions	0	0	2
Wrong referent	0	0	1
Total non-grammatical	49	67*	48
Grammatical			
Syntax	4	0	19
Omission of words	1	1	9
Prepositions	8	2	2
Word order	7	3	2
Addition of unnecessary words	1	1	8
Tenses	0	8	2
UDA	4	1	5
S+V agreement	0	0	8
V + preposition	3	4	0
Irregular plurals	1	2	2
Gerunds	1	1	2
3PS	0	2	2
Demonstratives	2	0	2
Quantifiers	1	1	0
Pronouns	1	1	0
Others	0	5	18
		verbs 2, pass. voice 2, adj. 1	verb form 3, rel. pron. 2, sing/plural 2, wrong word 2, modals 2, t. clause 2, quest construction, obj. pron. 5
Total grammatical	34	32	81**
Overall Accuracy	**83**	**99**	**129**

Note. *DIR*= direct group; *REF*= reformulation group, *SELF*= self-correction group, *GR*= grammatical, *N-GR*= non-grammatical

Table 31 summarises occurrence, number and type of features noticed reported on both NSs.

Table 31 – *Occurrence, amount of noticing and type of features noticed reported on noticing sheets.*

Observation	Noticing sheet-1 – composing stage	Noticing sheet-2 – feedback stage
OCURRENCE (or not) of both types of noticing	Noticing the hole (NS-1/ 40 out of 45) and noticing the gap (NS-2/ 40 out of 45) both occurred in the three experimental groups in the composing and feedback stages. Absence of noticing: blank noticing sheets also occurred in both stages. *Note:* blank NSs do not imply an absence of noticing (Schmidt, 1990)	
AMOUNT The three experimental groups differed in their amount of noticing in the composing and feedback stages	The number of overall reported features on NS-1 (52, 51, 38 = 141) was *LESS than* on NS-2	The number of overall reported features on NS-2 (83, 99, 129 = 311) was *MORE than* on NS-1
	If we focus on individual groups in the composing stage: DIR-G (52), REF-G (51), **SELF-G (38)**	If we focus on individual groups in the feedback stage: DIR-G (83), REF-G (99), **SELF-G (129)**
	SELF-G went: *FROM* noticing **the least** overall features	*TO* noticing **the most** overall features
TYPE Experimental groups (DIR, REF, SELF) looked at as a **single** group…	**N-GR** prevailed in both stages, i.e. N-GR features were reported as the: Most PFN in the composing stage (115 N-GR vs. 26 GR). Most FN in the feedback stage (164 N-GR vs. 157 GR).	
	Order: vocabulary, spelling, connectors, word repetition, content organization, punctuation, formal language.	**Order:** vocabulary, spelling, connectors, punctuation, content organization, formal language.
Experimental groups (DIR, REF, SELF) looked at as **individual** groups…	**N-GR** features still prevailed in the three experimental groups (as the most FN) considered together as a single group: 164 N-GR vs. 157-grammatical. BUT IN THE **FEEDBACK STAGE VARIATIONS AROSE**: **GR features prevailed in SELF-G** (81 GR) vs. DIR-G (34) and REF-G (32). Order: syntax, others, omission of words. A closer look shows a higher number of reports on text construction features in REF-G too.	

Note. NS = noticing sheet, *DIR* = direct group, *REF* = reformulation group, *SELF* = self-correction group, *PFN* = problematic features noticed, *FN* = features noticed, N-GR = Non-grammatical, GR = Grammatical.

11.3 Exit questionnaire results

The exit questionnaire was a supplementary data collection method used to enquire about learners' opinions and response to ECCs and learner-initiated noticing opportunities. Results are reported under each of the four questions: Learners' (Q1) impressions, advantages and disadvantages of the experienced ECC; (Q2**)** opinions on the usefulness of noticing sheet-1; (Q3) criteria to select what to focus on during feedback; (Q4) suggestions to make error correction more efficient.

Q1: Impressions about the experienced ECC

What is your impression (opinion or feeling) about the error correction technique offered to you? Advantages and disadvantages.

Direct error correction: Twelve positive, two negative, one blank. Twelve students (80% of participants) expressed positive opinions about DIR error correction. They emphasised its advantage in correcting all types of errors (grammar, spelling, lexis, style etc.) and in providing the correct version of their errors without ambiguity.

Students 3 and 6 (13.33%) had negative opinions. Student 3 argued that s/he learned more from indirect feedback as s/he enjoyed searching for his/her own errors. Student 6 said looking at his/her corrections was tedious and useless. "Despite this, there is improvement after some time," s/he added. Student 11 (6.66%) did not answer the question. Eight of the same participants above (53% of 15 students) made additional comments about writing and error correction in general. Students 1 and 12 described writing as a cognitively demanding skill. Students 2, 7, 13 and 14 remarked on the importance of making mistakes and the value of error correction. Students 8 and 9 stated their expectations from feedback: immediacy, explanation of every correction made and further work on mistakes.

Reformulation: Twelve positive, two 'balanced', one negative. Twelve students (80%) had positive opinions about reformulation. They remarked that reformulation was an informative technique and one never tried before. Its richness in vocabulary, formal language and content organization in addition to grammar accuracy was emphasised by six students (1, 9, 10, 11, 12, 14). "Reformulation helps students to express better and to include more appropriate words" (Student 1). "I see not only a correct version of my essay but different ways of expressing the same ideas" (Students 9 and 14). "It offers plenty of vocabulary and formal expressions" (Student 11). "We learn new vocabulary, understand grammar constructions and see what a fluent text is like" (Student 12).

Students 2, 4, 13 said reformulation had allowed them to look at their corrections from a freestanding perspective, which triggered meta-reflection. "Reformulation made me think about my writing. I identified new ways of building my sentences. It showed me more sophisticated ways of expressing my ideas" (Student 2). "The reformulated essay made me realize how much I still have to learn" (Student 4). "By comparing two versions of the same essay I identified errors I was completely unaware of. I observed a different way of assembling sentences. What I wrote was correct but my text was refined" (Student 13).

Students 8 and 15 remarked on the personal engagement in reading their own opinions rewritten in correct English. "This technique made me think about my text and about what I want to express: the communicative intention of my text, something we seldom deal with in class" (Student 8). "Corrections are easier to remember with this technique; this is what I wrote but it's improved" (Student 15).

Student 7 said reformulation increased his/her understanding of English syntax, text construction and punctuation: "The reformulated essay showed me the

real English syntax and grammar. My errors showed me I tend to structure my English writing with my L1 structure". Despite only one student saying this in the questionnaire, many others made this explicit on noticing sheet-2 (see common features of noticing sheet-2). "Something that attracted my attention was the use of punctuation … I realized how important punctuation is" (Student 8). Student 2 said s/he the reformulated version had made her/his text more engaging for the reader.

Students 3 and 6 (13.33%) whose opinions about reformulation were positive were classified as 'balanced'. This because although they appreciated the technique, they also said explanations of the reformulated parts were necessary. "We need to understand the reason for the reformulated ideas clearly" (Student 3). "Explanations are necessary to avoid repeating mistakes" (Student 6).

Student 5 (6.66%) complained about the ambiguity of reformulation, "I was completely confused with this technique. When I re-wrote my essay, I was unable to recall or to differentiate the correct from the incorrect or improved sentence. Thus, I formulated different ideas from those in the original essay."

Five of the same fifteen REF learners (33.33%) made additional comments on the strengths and flaws of reformulation. They are reported separately (a) because they were all individual opinions that might not be generalized and (b) to avoid overlapping with the total rates above. Student 9 mentioned reformulation might be suitable for only specific types or levels of learners "This technique is only suitable for people who give detailed attention to their feedback. At first sight I did not see my corrections." The risk of text appropriation was remarked on by Student 10, "Some of my ideas were changed", and Student 11, "The reformulator misunderstood some of my ideas." Despite being individual opinions, they are important, as they identify reformulation limitations found in previous studies (Cohen, 1983).

Self-correction: 10 positive, 3 balanced, 2 negative. Ten students (66.66 %) gave positive opinions about self-correction. They supported the idea that self-correction is not only important but necessary (Student 13) for learners to become independent and responsible for their learning, to pay attention to what they write, to recall and reinforce what they already know and to develop communicative strategies e.g. paraphrasing. "We become aware of our own weaknesses. Learning does not depend on the teacher but on how committed we are to learn" (Student 8).

Three students (1, 6, 14 = 20 %) were categorised as balanced. They all agreed self-correction was a good technique, though incomplete. Student 1 stressed that its success depended on learners' L2 knowledge. Student 6 explained he had not changed much in his self-correction as his/her problems were the same. However, s/he continued, "If I had known I was going to have a self-correction session, I would have looked for the words I needed." "Self-correction reinforced my English knowledge but there is much on writing essays I still need help with" (Student 14).

Two students (11 and 12 = 13.33 %) expressed themselves negatively. "I identified my errors but I did not know how to correct them. I preferred to change the sentence" (Student 11). "I wrote ideas the way I did because I think they were correct. I was unable to identify my own errors. Even if I were, I would not know how to correct them" (12).

Students 1 and 2 added having a time lapse before self-correction was convenient, as this had allowed them to find appropriate words and organize their ideas better. This suggestion was also mentioned on NS-2 (see common features of NS-2).

Table 32 – *Learners' impressions about different ECCs.*

Numbers represent number of students (% equivalent) out of fifteen students in each group.

ECC	Positive	Balanced	Negative	Blank
DIR	12 (80%)	- - -	2 (13.33 %)	1 (6.66%)
REF	12 (80%)	2 (13.33%)	1 (6.66%)	- - -
SELF	10 (66.66 %)	3 (20 %)	2 (13.33 %)	- - -

Note. ECCs = error correction conditions.

The responses in Table 32 show that most learners had positive views of the three ECCs. However, reformulation emerged as the most favoured technique fulfilling most of the learners' expectations. The two 'balanced' opinions were actually good but added contact with reformulators to make it more efficient. The richness of reformulation was perceived in its wide vocabulary, formal language, content organization and grammar in context.

I summarize learners' answers in this section by listing the characteristics of what could be considered learners' ideal feedback. Students want to be corrected comprehensively, unambiguously and immediately (DIR). Learners ask for feedback to be engaging, rich in input, cognitively challenging and reporting both strengths and weaknesses (REF). Promoting autonomy and self-correction is good, important and challenging; however, self-correction will only be effective if feedback is provided. Further work on their own corrections and more dialogue with people who correct their writing were also requested.

Q2: Opinion on the usefulness of NS-1

Did the identification of problems you had while writing (session 1) predispose/ prepare you to pay (more) attention to your feedback (session 2)? Why?

Thirty-five (77.77%) out of forty-five students answered affirmatively: 13 DIR, 12 REF, 10 SELF. Eight students (17.77%) answered negatively: 1 DIR, 2 REF,

5 SELF. There were two (4.44%) blank answers (DIR and REF groups). Affirmative answers showed identifying difficulties in the first essay made learners (a) revise their first essay and look for corrections in subsequent feedback. “When I received my essay back I looked for the connectors I had not recalled in the previous session” (Student 3 in REF-G); “I remember I did not how to say “oferta y demanda/ supply and demand”, so those were the first words I searched for in my corrected essay” (Student 12 in REF-G), (b) keep thinking about those problems even after class. “I kept thinking about my problem. When I saw the corrected essay the correction was better than I expected” (Student 2 in DIR-G). “I was attentive to solutions for my problems. Eventually, what I considered problems were actually correct sentences” (Student 10 in REF-G).

Q3: Criteria to select what to focus on

What were your criteria to select five things to attend to (noticing sheet-2)?

Q3 was replied to exclusively by the DIR and REF groups. SELF-G also answered the question. However, lack of external explicit feedback in this group led to learners’ confusion, which resulted in repetition of their answers to other questions. I decided not to include the answers from this group. Seven criteria were identified by the DIR and REF groups. As six students (one in DIR-G and five in REF-G) mentioned up to three criteria, I got 40 (16 in DIR-G + 24 in REF-G) instead of 30 answers. Table *33* presents the outcomes considering 40 as the total. The number of cancelled (unclear or deviating) answers for this question was higher than in other questions. However, with the aim of distinguishing the criteria guiding learners’ attention, I rated the number of deviating answers and included it in the total. The results suggest “errors that are particular to each student as the top criterion”. This category refers to learners’ reports referring to specific problems they think they have.

Examples of learners' answers included in this category are: "Things I know I particularly have problems with", "Syntax and grammar are priorities for me", "Wrong use of tenses with similar meanings is common in all my writing tasks." Examples of language features mentioned in this category were: tenses, passive voice, prepositions, possessive 's', irregular plurals, conditionals, unnecessary words and 3rd person singular. Frequency was the second criterion; it seems logical that recurrence influences noticing. Criteria 3 and 5 confirm Nassaji's (2010) claim that the best moment to promote noticing is when learners have trouble making their message precise. The need for synonyms and referents (fourth criterion) to avoid word repetition is new to me and came up on the noticing sheets too. Novel language came up in noticing sheet-2 analyses. Formal language was the last criterion mentioned.

Table 33 – *Criteria to 'attend to' reported by experimental groups.*

Criteria	**%**	**No. students per group**
Errors that were particularly theirs (tenses mainly)	25	10 (3 DIR, 7 REF)
Cancelled answers (deviating, unclear answers)	17.5	7 (5 DIR, 2 REF)
Most repetitive errors	15	6 (4 DIR, 2 REF)
Words or structures essential to express what they meant to say or to understand his/her correction	12.5	5 (1 DIR, 4 REF)
Synonyms or referents provided in the feedback for which they had repeated the same word(s)	12.5	5 (2 DIR, 3 REF)
Sentences, words or structures they struggled to formulate	5	2 (REF)
New expressions or structures	5	2 (1 REF)
Blank answers	5	2 (1 DIR, 1 REF)
Formal language	2.5	1 (1 REF)

Note. *DIR* = direct group, *REF* = reformulation group

Q4: Suggestions to make EC more efficient

How would you advise teachers to make error correction more efficient for you?

Three students (6.66%): Student 11-DIR, Student 5-REF, Student 15-SELF did not reply. Seven students (15.55%): four in DIR-G (Students 6, 9, 10, 14), three in

REF-G (Students 7, 8, 11) asked for an explanation for every correction made. Unless they understand the reasons for their errors they cannot avoid making them again, they said. Seven students (15.55%): two in DIR-G (Students 12, 13), two in REF-G (Students 1 and 10), three in SELF-G (Students 2, 5, 10) suggested including a summary of their errors at the end of each piece of writing. They said feedback had to be organized so that teachers could let them know what their most frequent errors were, what caused them and provide examples of how to solve them. Three students (6.66%): Student 4 in DIR-G, Student 15 in REF-G, Student 14 in SELF-G proposed writing shorter but more frequent essays to help their little writing practice in class. Five different students (11.1%) made individual requests, one each (2.22%): Student 1 in SELF-G requested the correction of all types of errors, Student 15 in DIR-G asked to use noticing sheet-1 more often, Student 3 in REF-G asked to use REF regularly, Student 9 in SELF-G said using SELF-correction was important, Student 2 in DIR-G requested that original wrong sentences not be deleted because looking at them was important to understand corrections. Six different students (13.32 %), two students (4.44%) for each request, suggested: more teacher-student dialogue while providing feedback (Student 3 in DIR-G, Student 12 in SELF-G); feedback to be immediate (Student 9 in REF-G, Student 13 in SELF-G); including practice of their corrected problems as part of the feedback (Students 3 and 8 in SELF-G). Fourteen students (31.11%): four in DIR (Students 1, 5, 7, 8): six in REF (Students 2, 4, 6, 12, 13, 14); four in SELF (Students 4, 6, 7, 11) made unrealistic recommendations, such as hiring teaching assistants to get essays corrected as soon as possible, providing personalized feedback, providing resources and materials for them to practise. The answers above complement the “learners’ ideal feedback” list I started in question one.

SECTION 5: DISCUSSION AND CONCLUSION

Section 5 ‘Discussion and Conclusion’ is composed of 4 chapters. Chapter 12 answers the research questions and discusses the findings for each question. Chapter 13 discusses the findings from the questionnaire, a supplementary data collection method. Chapter 14 examines additional findings in the study. In Chapter 15, reflections on the theoretical and pedagogical implications of the study are presented. Chapter 16 states the limitations of the study. Finally, Chapter 17 closes with a general summary and outlines directions for future studies.

Chapter 12: Answers to Research Questions

12.1 Research Question 1

RQ-1: What are the effects (if any) of different comprehensive (eventually replaced by semi-comprehensive) ECCs (DIR, REF and SELF) on learner-initiated noticing at two stages (composing and feedback) of the writing process?

1.a - Do different comprehensive (eventually replaced by semi-comprehensive) ECCs (DIR, REF and SELF) lead to learner-initiated noticing in the composing and feedback stages? i.e. The occurrence of noticing reports.

1.b - If so, how much noticing and what error types (grammatical, non-grammatical) do learners pay attention to by themselves in each ECC at the composing and feedback stages, i.e. The number of noticing reports and error types reported as noticed.

Results from the NSs indicate that different ECCs had different effects on learner-initiated noticing in the composing and feedback stages. The occurrence of learner-initiated noticing was common to all ECCs; all tested ECCs led to learner-initiated noticing. Differences arose in the number and the type of features noticed reported in each ECC.

12.1.1 Occurrence of both types of noticing

Both noticing the hole (NS-1) and noticing the gap (NS-2) occurred in all three experimental groups in the composing and feedback stages. An absence of noticing occurred in both stages too. Five learners in the composing and five in the feedback stages did not complete NSs. Blank NSs, however, did not imply an absence of noticing as some experiences are hard to express (Schmidt, 1990). Minimal instructions and the opportunity to leave NSs blank (if learners did not want to complete them) were hazardous (as no emerging data was possible) but necessary to guarantee learner-initiated noticing. Students' reports on the NSs showed the above

decisions worked well. Learners examined their writing (NS-1) and WCF (NS-2) and provided the researcher with insights into their interests, challenges and views.

12.1.2 Amount of reported noticing

The three experimental groups differed in their amount of reported noticing in the composing and feedback stages. The number of overall noticed features reported in the feedback stage (311) was higher than in the composing stage (141). These results were determined by SELF-G, whose results were thought-provoking, as it was the group where no external explicit feedback was provided that reported the most noticing. SELF-G went from reporting the least overall features (38) in the composing stage to reporting the most overall features (129) in the feedback stage, nearly 30 per cent more than REF-G and 20 per cent more than DIR-G.

Differences in the amount of reported noticing among the groups in the composing stage may be explained by individual learners' differences in each group. Lack of familiarity with NSs and learners' habit of making teachers' responsible for correction are possible explanations too. Regarding the increased amount of reported noticing in in the experimental groups in the feedback stage, this was to some extent expected as a result of internal (SELF) and external input via feedback (DIR, REF). What is key here is the noticeable increase in reported noticing in SELF-G which highlights the potential of self-provided or self-initiated feedback to promote noticing. The absence of external explicit feedback and instructions on NS-2 in this group revealed this potential.

12.1.3 Error types reported as noticed

Non-grammatical error types prevailed in learners' reports in the three experimental groups in the composing and feedback stages, i.e. non-grammatical features were reported as the most problematic features noticed in the composing

stage (115 non-grammatical vs. 26 grammatical) and the features most noticed in the feedback stage (164 non-grammatical vs. 157 grammatical). The main problematic features according to learners' reports were not grammatical. Learners' needs while composing were mainly in lexis, spelling, lack of synonyms, lack of connecting devices and non-grammatical features in general. The predominance of non-grammatical features in the composing stage might be explained by learners' prioritising meaning, which is to be expected in this stage. Learners' struggles to communicate their ideas made them aware of vocabulary they did not know or could not remember, constant repetition of the few words they could recall, a shortage of formal language and so on. In order of frequency, lack of lexis was the single most reported problematic feature, followed by spelling, lack of connectors, word repetition, content organization, unknown punctuation rules and limited formal language.

Lexis and spelling were the only two error types, of the seven analysed in this study, that learners considered problematic. Third person singular (3PS), for instance, was only mentioned by one learner in SELF-G. Grammatical features were much less reported in the DIR and SELF groups (five and four, respectively) in NS-1. Error frequency analysis and problematic features noticed matched with non-grammatical error types only: lack of lexis and wrong spelling. The remaining five selected errors in this study were grammatical which, as I said before, are less important for learners. Considering order of frequency, almost the same non-grammatical features were reported in the composing and feedback stages, see Table 34, below.

Table 34 – *Frequency order (top to bottom) of non-grammatical features reported in the composing and feedback stages.*

This table shows that the order of reported features noticed was almost the same in both stages in all groups.

Non-grammatical features	
Composing	**Feedback**
Reported problematic features noticed	Reported features noticed
Lexis	Lexis
Spelling	Spelling
Connectors	Connectors
Word repetition	
	Punctuation
Content organization	Content organization
Punctuation	
Formal language	Formal language

Regarding reported error types in the feedback stage (NS-2), however, some variations arose. Although non-grammatical features still prevailed in the DIR (49 grammatical vs. 34 non-grammatical) and REF (67 grammatical vs. 32 non-grammatical) groups, and in the three experimental groups considered together as a single group (164 non-grammatical vs. 157 grammatical), features noticed reported by learners in SELF ECC were mostly grammatical, with almost double (81) the number of non-grammatical errors in DIR (34) and REF (32) in the same stage.

SELF-G was in fact the only group where attention was paid to grammatical features. These results may contribute to the unresolved issue of how different types of WCF influence noticing and uptake (Santos et al., 2010). The lack of input via feedback in SELF-G might explain learners' attention to form. Knowledge of lexis, punctuation, connectors and formal language, on the other hand, might be less or non-existent in learners' background explicit knowledge; thus, they may require input via feedback or other sources (e.g. dictionaries, instruction). One of the students in SELF-G said, "If I had known I was going to be given the same text in the following class, I would have looked for the words I needed." Attention to grammatical features is, however, also evidence of learners' ownership of explicit grammatical knowledge that

allows them to identify their errors. Student 13-DIR said, “I should be more careful with pronouns, articles etc. Although I know how to use them, I used them incorrectly.” Among grammatical error types reported, syntax was also the highest reported feature in SELF-G; this was mainly because learners modified full sentences during self-correction.

Results of features noticed in REF-G NS-2 support Hanaoka’s (2007) and Santos et al.’s (2010) studies’ findings. Both investigations showed that learners’ noticing of lexical features was higher in REF-G. In Santos et al. (2010), learners reported that incorporating discourse features from their reformulated text was more difficult than lexical features. In the present study, REF-G also reported the highest number of lexical items in the feedback stage: REF (32), DIR (25), SELF (18).

A closer look at NS-2 in REF-G (after lexis and spelling as first and second top errors, respectively) reveals that learners in this ECC reported the highest numbers of text construction features: connectors: REF (11), DIR (5), SELF (3); punctuation: REF (7), DIR (2), SELF (1); content organization: REF (5), DIR (1), SELF (1); formal language: REF (4), DIR (1), SELF (0). These features might have attracted learners’ attention in REF-G because these were new input for them. Learners are likely to consider text construction features secondary in writing development as most EFL classrooms are dominated by focus on form (Zamel, 1985; Lee, 2011; Evans et al., 2010; Giraldo de Londoño and Perry, 2008).

12.2 Research Question 2

RQ-2: What are the effects (if any) of the above noticing-supported ECCs at the feedback stage on learners' writing accuracy in the rewriting and new writing stages?

2.a - Does the input provided by the above noticing-supported ECCs at the feedback stage have any effects on learners' writing accuracy in the rewriting and new writing stages? i.e. The occurrence of effects of noticing-supported ECCs in the rewriting and new writing stages.

2.b - If so, what error types (grammatical or non-grammatical) are more amenable to correction in different noticing-supported ECCs? i.e. The error types learners corrected by themselves in different ECCs.

The results from descriptive statistics revealed that there were positive effects from noticing-supported ECCs in the feedback stage on learners' writing accuracy in the rewriting and new writing stages. All groups showed overall accuracy improvements from essay 1 (composing) to essay 2 (rewriting). The reason for this immediate accuracy improvement might be attributed to the treatment (error correction and noticing opportunities). Whether it was the error condition, the noticing opportunities or a combination of both, what led to improvement could not be determined at this stage. These descriptive statistics outcomes support findings in previous studies regarding the positive effects of different ECCs in a post- or immediate writing task, i.e. WCF as a revising tool (e.g. Ashwell, 2000; Fathman & Whalley, 1990; Ferris & Roberts, 2001). What is new in my results is that they come from a less common task type: an open uncontrolled learner-produced task, with semi-comprehensive error correction. DIR, REF and SELF ECCs showed positive effects in an immediate test. However, Santos et al. (2010) suggest immediate improvement cannot be considered acquisition. In fact, it is possible that recalling played a role in immediate improvement. Except for the control group, learners in all the ECCs were engaged in processing their feedback for at least an hour. Thus, it is natural that, even if unintentionally, learners remembered input they had just been processing.

Descriptive statistics also showed this immediate accuracy improvement was partially retained in T3 (new test or delayed post-test) in the two groups receiving external explicit feedback (DIR and REF groups). From test 2 to test 3, the DIR and REF groups increased the error mean that had decreased for both groups from test 1 to test 2. The increased error means of both groups in test 3, however, were still below (DIR-G) or equal to (REF-G) their error mean in test 1. This may suggest that some of the accuracy improvement in test 2 was retained until test 3. This finding (from descriptive statistics only) also supports previous studies' (Bitchener, 2008; Bitchener & Knoch, 2008; Bitchener & Knoch, 2009a; Bitchener & Knoch, 2009b; Sheen, 2007; Storch, 2009) findings of accuracy improvements being retained in new writing or a delayed post-test.

Contrariwise, the numbers of errors in SELF-G and CONTROL-G from test 2 to test 3 decreased. The behaviour of these two groups will be discussed below, under results with statistical significance. Although descriptive statistics results showed a positive impact from all noticing-supported ECCs on overall accuracy improvement in the rewriting and to a lesser extent the new writing stages, statistical significance did not support these results. Statistical significance across the three tests was exclusive to: (a) overall and non-grammatical accuracy in DIR-G and (b) non-grammatical accuracy in SELF-G. Concerning the selected seven error types, a series of ANOVAs revealed that except for spelling in the DIR and CONTROL groups, the error type did not influence accuracy performance.

The effects of noticing-supported ECCs on accuracy improvement supported by inferential statistics are summarized and further discussed in the following findings.

12.2.2 No effects from noticing-supported ECCs on GR accuracy

Statistical significance being found exclusively for DIR-G (in overall and non-grammatical accuracy) and SELF-G (in non-grammatical accuracy) means little if any effect of noticing-supported ECCs on grammar accuracy improvement. The prevalence of non-grammatical over grammatical features in significant accuracy improvement results might lend support to Truscott's claim for the inefficacy of WCF for grammatical features in the three tested noticed-supported ECCs. There were no effects in terms of grammatical accuracy improvement, even in descriptive statistics. This outcome contradicts Van Beuningen et al.´s (2012) conclusion about the efficacy of comprehensive WCF (specifically direct comprehensive WCF) for-improvement of grammatical errors. Support for Truscott's claim in this study may derive from the uncontrolled open writing task, and the comprehensive error correction learners received. The latter refers to the fact that learners received comprehensive EC of their essays; semi-comprehensive EC was used only for essay accuracy analyses. Research evidence opposing Truscott comes mostly from focused WCF studies using controlled writing tasks. As long as evidence for the benefits of WCF comes from focused rather than unfocused experimental studies, it will be difficult to refute Truscott's claims (Ellis et al., 2008). Among the five studies enquiring into unfocused WCF, my results support Truscott and Hsu's (2008) conclusion regarding the lack of effects of unfocused WCF on grammatical accuracy. The error categories targeted in Truscott and Hsu were mainly grammatical (orthographical and lexical errors were not corrected) because Truscott's claim opposing the efficacy of WCF refers exclusively to grammatical features. Improvement in only non-grammatical accuracy also supports Van Patten's (1994) claim that, in language processing, content or meaning

rather than linguistic features or form is processed first, accuracy is secondary for comprehension, and for WCF at the composing stage too, I would add.

12.2.3 Effects of noticing-supported DIR EC on N-GR accuracy

Statistical significance across the three tests was found in DIR-G for overall and non-grammatical accuracy, the latter influencing the former. This finding agrees with the conclusions of previous studies regarding the efficacy of DIR-WCF in post and delayed-post writing tasks (Chandler, 2003), on the advantages of DIR over REF (Sachs & Polio, 2007; Santos et al., 2010) ECs. The reasons for the success of DIR-ECC for non-grammatical accuracy may be various: (a) corrections are more explicit in DIR-ECC than in REF and SELF-ECCs. Manchón (2011b) and Sheen (2010) regard the degree of explicitness of WCF as one of the most influential factors for its success; (b) locating errors and their corresponding corrections is easier and less confusing in DIR than in REF and SELF-ECCs. Santos et al. (2010) explain that corrections in DIR-ECC are more salient than in any other type of feedback; (c) the number of errors or changes in DIR-ECC is also less than in REF-ECC; (d) learners' familiarity with this technique is likely to have an effect too; (e) DIR-ECC meets learners' expectations of clear and direct correction of every single error; (f) computer-mediated correction might have overcome the messy corrections of DIR-EC using pen and paper.

12.2.4 Effects of SELF-EC on N-GR accuracy

The fact that SELF-G (with only noticing opportunities and self-provided or self-initiated feedback) was the only experimental group that improved its non-grammatical accuracy across the three essays suggests that accuracy improvement might not be a consequence of ECCs but of noticing opportunities. Although supported by inferential statistics (statistical significance was found for non-

grammatical accuracy improvement in SELF-G between T1 and T3), this conclusion lacks strength due to the absence of a fourth experimental group (one receiving only ECC) in the design. Self-correction is a frequently recommended correction technique in the SLA literature and language teaching practice. However, evidence for its efficiency is rarely provided. The significant statistical support for non-grammatical improvement in SELF-ECC refutes Van Beuningen et al.'s (2012) conclusion that WCF is more beneficial for learning than self-correction and sheer writing practice.

12.2.5 Effects of sheer practice on N-GR accuracy

The spelling accuracy improvement found in DIR-G (with external explicit feedback) and CONTROL-G (without any type of feedback) is interesting. Spelling improvement in DIR-G vs. REF-G (both receiving external explicit feedback) could be explained by the richness of input in the REF condition that attracted learners' attention to other more relevant features, rather than spelling. However, spelling being significant in CONTROL-G is different. That CONTROL-G managed to improve spelling accuracy without feedback and without noticing opportunities suggests that mere practice might be enough to improve this specific error type. Truscott (1996) claims learners' time and effort would be more productively spent on writing practice. Sheen et al.'s (2009) study did not find evidence for the efficacy of sheer practice. Considering spelling only, this study may contribute some evidence. This finding, already mentioned in answer to RQ-2, also reinforces task proponents' claim for the usefulness of task repetition (Bygate, Lynch and Maclean, 2001). Recalling Van Patten's (1994) suggestion that different language aspects require different amounts of attention, the results suggest that spelling may only require attention at the 'noticing level' to see improvement. If spelling is one of the most treatable errors, teachers

might now have justified reasons to pass the responsibility for spelling correction to learners themselves.

12.2.6 Excluding spelling, no error type effects in writing accuracy

Moving on to the results for the seven selected error types, statistical significance across the three tests was found for spelling in the DIR and CONTROL groups. These results may suggest two things. First, except for spelling in DIR and CONTROL groups, errors in writing accuracy were not influenced by the tested ECCs. Second, learners were able to correct their spelling errors with (DIR-G) and without (CONTROL-G) feedback. The results support Truscott's (2007) statement that "spelling errors are among the most correctable error types because they are relatively simple and can be treated as discrete items" (p. 258). In his 12-error category study, Lalande (1982) also found an 83 per cent improvement for orthographic errors. Spelling has seldom been explored in WCF research. The high frequency of spelling errors that emerged in learners' essays in this study, however, deserves attention. In my view, spelling accuracy has been undervalued in L2 writing, despite being a feature that contributes greatly to L2 writing accuracy. The results in the present study suggest not only that spelling is the most frequent error in learners' writing but also one most amenable to correction (Truscott, 2007; Lalande, 1982). In the search for writing accuracy, all features (grammatical and non-grammatical) should be addressed.

To sum up the discussion on RQs, different ECCs might influence the occurrence, and amount of noticing as well as the error types noticed by learners. ANOVA accuracy tests on essay writing showed that DIR-ECC resulted in the most efficient ECC for overall and non-grammatical accuracy improvement across the three tests (pre-test, post-test and delayed post-test). Accuracy improvement was especially

good in the rewriting stage (test 2), where both lexis and spelling accuracy improved the most. REF-ECC elicited the most attention to non-grammatical features, particularly lexis and text construction features: connectors, punctuation, content organization and formal language. Learners in this group said they did not know these features were as important as grammar. This suggests the REF-ECC allowed learners to gain in awareness of writing beyond grammatical error-free sentences. SELF-ECC was the one that elicited the most attention to form or grammatical features. CONTROL-G showed that sheer practice does play a role in spelling accuracy improvement; statistical significance supported this finding.

Table 35 contrasts error types reported as noticed in the composing and feedback stages vs. improved error types. Except for SELF-G on NS-2, non-grammatical error types were the most reported as noticed on NS-1 and NS-2. Regarding improved error types, statistical significance was only found for non-grammatical error types in (a) DIR-G (N-GR errors determined the overall error type accuracy); (b) SELF-G and CONTROL-G (determined in spelling). This means that non-grammatical errors were the only ones to improve in accuracy.

Table 35 – *Error types reported as noticed vs. error types showing accuracy improvement*.

This table shows N-GR error types were the ones most reported as noticed in the composing and feedback stages. Similarly, accuracy improvement emerged for N-GR errors in the DIR, SELF and CONTROL groups.

<table>
<tr><th colspan="3">Error types reported as noticed</th><th colspan="3">Error types showing accuracy improvement</th></tr>
<tr><th>Experimental ECC</th><th>Composing
(NS-1)</th><th>Feedback
(NS-2)</th><th>ECC</th><th>Overall, GR, N-GR</th><th>Error type</th></tr>
<tr><td>Considered together as a single group:
DIR, REF and SELF</td><td colspan="2">N-GR error types prevailed in both stages</td><td>DIR</td><td>Overall</td><td></td></tr>
<tr><td>Considered separately as individual groups:
DIR and REF</td><td colspan="2">N-GR error types still prevailed in both stages in DIR and REF groups. However, …</td><td>DIR</td><td>N-GR</td><td>Spelling</td></tr>
<tr><td rowspan="2">Ditto: SELF</td><td rowspan="2"></td><td rowspan="2">… GR error types prevailed in the SELF group in the feedback stage</td><td>SELF</td><td>N-GR
(T1-T3)</td><td></td></tr>
<tr><td>CONTROL</td><td></td><td>Spelling</td></tr>
</table>

Note: ECC= error correction condition, *NS*= noticing sheet, *GR*= grammatical, *N-GR*= non-grammatical. *DIR* = Direct group, *REF* = Reformulation group, *SELF* = Self-correction group

Table 35 raises some questions. Most of them have been discussed throughout this chapter. Still, a recapitulation of the answers is presented below.

If DIR-ECC is much criticised for being overwhelming for teacher and students, how do we explain its efficacy in accuracy improvement? Empirical studies exploring the feedback stage in WCF (e.g. Sheen, 2010) identify the degree of explicitness of feedback as a feature that influences short-term learning in immediate tests. Among the ECCs tested in this study, feedback was most explicit in DIR-ECC.

If REF-ECC is assessed as being the richest and most engaging WCF technique, why were these advantages not reflected in learners' writing accuracy improvement? Feedback in REF-ECC is less explicit than in DIR-ECC. Thus, in agreement with previous studies (Sachs & Polio, 2007; Santos et al., 2010), my results support DIR-ECC as the most efficient way to lead to immediate short-term writing accuracy. Nevertheless, this does not imply that DIR-ECC will be more beneficial in the long term too. The effects of engagement with feedback in REF-ECC require long-term measures and/or individual case studies.

What does N-GR accuracy improvement in the SELF and CONTROL ECCs mean? N-GR accuracy improvement in the SELF and CONTROL ECCs apparently supports the small effect of teacher-provided error correction. Both, the SELF and CONTROL groups received no external explicit error correction. Despite this, the learners in those groups improved their non-grammatical accuracy. It is important to note that non-grammatical accuracy improvement was determined by spelling, a treatable error type. A better interpretation of the results might be that, after the first composing stage, noticing opportunities should be added to the writing process. Two reasons justify this; first, the above results show that after their first composing, learners were still able to improve their written accuracy by themselves, even if it was

only the accuracy of non-grammatical treatable errors. The second reason refers to attention being limited, which obliges learners to distribute their time across task stages; in the composing stage, learners concentrate on conveying meaning, so, another stage is needed to address their attention to form.

How do we explain students' outstanding attention to grammatical error types in SELF-G in the feedback stage? Differences in the amount of reported noticing among the groups from the composing to the feedback stage were, to some extent, expected as a result of internal (SELF) and external input via feedback (DIR, REF). However, the results also suggest various things. First, learners' small amount of reported noticing in the composing stage suggests their lack of familiarity with noticing opportunities and their habit of making teachers' responsible for correction. Second, the noticeable increase in reported noticing in SELF-G suggests the potential of self-provided or self-initiated feedback to promote noticing. The absence of external explicit feedback and of detailed instructions on NS-2 in this group revealed this potential. Third, learners' attention to form, in SELF-G might be explained by the lack of feedback in this group. However, this in turn suggests learners' ownership of explicit grammatical knowledge that allows them to identify their errors. This, added to the above recognition of learners' capacity to improve the accuracy of non-grammatical treatable errors, expands learners' capacity for accuracy improvement to account for grammar errors too, at least those learners who are familiar with them, i.e. mistakes due to processing inefficiency to cope with the demands of the task. Knowledge of lexis, punctuation, connectors and formal language might be less or non-existent in learners' background explicit knowledge; thus, they may require input via feedback or other sources or formal explicit instruction. Finally, the task stage comes up in this study as a variable that influences what and how much learners

notice. As noticing capacity is limited, in the composing stage learners concentrate their attention on conveying meaning (non-grammatical features), whereas in the delayed self-correction[18] stage, when meaning has been dealt with, learners' attention can be addressed to form. This supports Swain and Lapkin's (1995) findings. Their participants paid more attention to lexis in the drafting stage, whereas grammar received more attention during the editing phase of composing.

What does the prevalence of N-GR (specifically spelling) over GR error types in reported noticing and accuracy improvement suggest? The prevalence of N-GR (specifically spelling) over GR error types in reported noticing and accuracy improvement supports Truscott's claim for the small effect of grammar correction in WCF, at least when referring to open learner-produced writing tasks with unfocused EC. The lack of effect of grammar correction is however, only valid when such correction is provided for learner's first composition. Considering learners' attention to grammar in SELF and the non-grammatical improvement of written accuracy in CONTROL, it might be possible that we, as researchers, have looked for evidence for the value of grammar correction in the wrong writing stage. Grammar correction may be effective, but finding evidence for this requires treating writing as a process rather than a product, as most experimental studies do. The view of writing as a product reduces the role of WCF to information transmission. Hence, future research could explore the effects of grammar correction provided in the editing stage.

[18] Delayed self-correction is task stage I propose as an opportunity for learners' self-initiated noticing from their delayed self-correction or delayed self-provided feedback. It is explained in detail in the "Additional findings of the study" section.

Chapter 13: Questionnaire Findings

Information from the exit questionnaires was only complementary. I used it to supplement information data collected using other instruments. Retrospective questions are criticized for being an incomplete reflection of learning processing (Matsumoto, 1993). However, gathered with care, learners' answers also became a rich source of information about learners' views of L2 learning, learning expectations, interests and goals that would have not been available if retrospective open-ended questions had not been used. Despite the methodological disadvantages of delivering the questionnaire at the end of the experiment, I find having the information obtained more beneficial than no having information at all.

13.1 Reformulation and SELF-EC

Most learners had positive views of the three ECCs. However, their answers also revealed that individual differences, such as learning style, view of language and L2 level, made learners see advantages and disadvantages in all types of ECCs. Learners' opinions about DIR-ECC were quantitatively equal to REF (both 80 per cent acceptance). I present REF as the most engaging technique because of the amount of qualitative information learners added to their answers. Opinions about the DIR-ECC were limited to indicating positive or negative views with no additional information. There may be three reasons for this: (a) learners are used to the DIR EC technique; (b) DIR EC meets learners' expectations of clear and direct correction of every single error; (c) most participants had never tried REF therefore, they could not judge what they did not know.

Reformulation emerged as the ECC fulfilling most of the learners' expectations: innovative, enriching and stimulating. Its strengths were perceived in its

extensive vocabulary, formal language, content organization and grammar in context. Reading a correct version of their own writing and content might be its greatest advantage, not only because it was engaging for them to read their own text, but also because reformulation implicitly emphasised learners' communicative strengths rather than their flaws. Reading their reformulated text allowed learners to (a) focus their attention on other features beyond grammatical and non-grammatical accuracy: cultural aspects in essay writing, well-structured text, punctuation, formal refined language; (b) gain awareness of writing as something more than grammatical accuracy. For example, "I was surprised to realize I do not only have English errors but writing errors. This alarms me. I will consider taking a writing and punctuation course" (Student 12-REF). Learners' awareness of the complexity of writing also emerged in the pilot study and became one the most rewarding gains from the use of reformulation. Learners differentiated reformulation from traditional DIR unfocused EC. They called the former authentic feedback. Student 1 in the pilot study, where participants experienced both types of feedback, said: "With traditional error correction I understand my problems are with the corrections made to my writing but I do not get to see what I see with reformulation. My corrected essay is still far from being native-like." Important for the future implementation of this technique, however, are: (a) the need for explanations of the reformulated bits; (b) noting that although REF worked well with this proficiency level but (as shown in the pilot study), it might not succeed with lower L2 levels.

Regarding SELF-ECC, opinions were surprising, and rewarding too. Eighty-five per cent of learners had positive opinions about it. Learners appreciated being given the chance to correct themselves. Their amazement at their ability to self-correct was expressed in phrases like:

‘I was surprised I was able to correct many errors by myself’ (Student 4). ‘Translations that seemed correct when I wrote them looked weird when I revised my essay’ (Student 9).

13.2 Support for noticing the hole

The inclusion of NS-1 in this design aimed to find out whether and how much ‘noticing the hole’ occurred in WCF. Different from Hanaoka (2007), whose objectives required tracing problematic features noticed in NS-1 and NS-2, the role of NS-1 in this study was assessed considering only learners’ answer to exit question 2. Learners indicated that prompting noticing of their writing problems in the composing stage (NS-1) had predisposed them to attend to their feedback. Thirty-five students (77.8%) out of forty-five confirmed this. Learners claimed identifying problematic features in essay 1 made them think about those problems even after the session. These features, they said, were the ones they looked for when receiving feedback. The methodological drawbacks of working at the level of perception only and of delivering this question at the end of the experiment (when learners might not recall their noticing experience) have to be considered when valuing these results. Learners’ positive responses, however, are important for further exploration of the inclusion of this type of noticing in the classroom.

13.3 Attention criteria

The learners’ attention criteria reported in this section might reinforce inferences or uncover data teachers and researchers may already have about factors influencing noticing. The top criterion was (a) attention to participants’ personal errors, i.e. errors particular to each one of them. This suggests learners have different linguistic needs they might be aware of, needs they need to address or monitor more. (b) Essential words or structures to express their thoughts or understand their

corrections were another criterion. Learners seemed to focus on transmitting content rather than on writing accurate sentences in the composing stage. (c) As important as the previous criterion were synonyms and/or referents, provided in their feedback, for which learners had constantly repeated the same word(s). Learners' worry about lack of synonyms leading them to constantly repeat the same words was also stated on the NSs. This might be new information for teachers and may deserve attention in instruction. (d) Recurrent errors, another criterion mentioned, are natural to be noticed. Attention to sentences, words or structures that learners struggled to formulate supports Nassaji's (2010) opinion on this stage being the best moment to provide feedback. (e) Interest in new expressions and structure(s) was also mentioned. Other interesting criteria were brought up in the pilot study, worth mentioning here are sentences that were very different from learners' original ones, especially those that had made their essays clearer and more coherent, or errors that could hinder clear communication.

13.4 Making WCF efficient

Learners' ideal feedback can be built from learners' answers to this question. Learners expect WCF to be engaging, rich in input, cognitively challenging and to report strengths and weaknesses. Learners signalled that promoting autonomy and self-correction is essential; however, they also called for more teacher-learner dialogue, writing shorter but more frequent essays, and using NSs and reformulation in their instruction. Learners' answers implicitly reflected some of their beliefs about EC: (a) requests for explanations of every correction made, for example, revealed their faith in comprehensive EC; (b) requests for teachers to include a summary of their errors at the end of their writing reveal, on the one hand, learners' view of teachers as being responsible for their language processing but, on the other hand, learners' lack

of awareness that writing such summaries themselves will allow them to better assimilate language. (c) Finally, learners' requests for short but more frequent writing practice might be pedagogically valid and useful information for teachers.

Chapter 14: Additional Findings of the Study

14.1 Individual learners' differences

The noticing sheets succeeded not only in encouraging learner-initiated noticing in a non-disturbing way and in providing information about what learners give importance to while writing and receiving feedback, but also in eliciting additional information about learners' individual differences. These refer to the idea that every person has a distinctive mixture of characteristics (personality, motivation, learning style etc.) that determine their learning outcomes (Murphy & Falout, 2010). Murphy and Falout (2010) explain that traditional and recent approaches to learners' individual differences currently discuss whether individual differences are unchangeable or malleable and socially interdependent traits.

By analysing their written essays and feedback freely and independently, learners made it possible to identify individual differences vis-à-vis introspection and some of their approaches to feedback: their writing strategies; their views on writing, error correction and language learning; some of their cognitive and metacognitive processes. To a lesser extent but also contributing to the findings for individual differences were answers to the questionnaire. Learners´ opinions about different ECCs disclosed different learning styles and probably different L2 proficiency levels. Their answers to Q4 revealed: (a) their reliance on comprehensive EC (they asked for correction and explanation of every single error); (b) their lack of learning autonomy (they would like teachers to write a summary of their weaknesses for them); (c) their conception of teachers as responsible for their learning and language processing; (d) their goals (learners worry about lack of synonyms and referents as part of their writing skill).

Concerning *writing strategies*, for instance, learners reported on the NSs: paraphrasing, describing unknown vocabulary and avoidance of not-yet-mastered structures as some of the strategies they use to communicate in writing. For example: "I have to give less information when I write and increase my vocabulary. When I do not know a word, I explain it and this makes my essay longer and confusing" (Student 10-DIR).

Evidence to support Truscott's claim for learners' avoidance of unmastered linguistic features and simplification of language was actually found on NS-1. Learners admitted that some non-grammatical problems, such as failure to start paragraphs or link ideas (Student 5-REF), lack of vocabulary (Student 15-REF), lack of synonyms (Student 6-REF; Student 13-REF) and spelling uncertainties (Student 15-REF), had led them to: (a) omit or modify what they were trying to say; (b) write incomplete sentences; (c) constantly repeat words.

Learners' view of [good] writing as producing grammatically correct sentences was also unanimous, so were their views of error correction and language learning; the latter focused on grammatical accuracy too. For example: "Most of my corrections were in vocabulary. I was worried about grammar. I need to learn more connectors: however, but…" (Student 14-DIR).

What, how much and how deep learners reported, the layout they used to report (charts, lists, plain sentences or bullet points), questions they raised and calls for help were evidence of cognitive processes and feedback processing. NS-2 also led to metacognitive reflection. Some learners did not only categorise errors as adverbs, adjectives, idioms, lack of connectors, discourse markers and lexis, but frequently reported actions to overcome their errors, e.g. using reformulation more often, copying texts to improve spelling or further reading.

Learners' beliefs, attitudes, aptitudes, personalities and goals, Storch and Wigglesworth (2010) point out, "influence not only the strategies learners adopt in dealing with the feedback received […] but also their willingness to accept the feedback and the likelihood of retaining it" (p. 328). Learners' individual differences are intrinsic to the learning process. Every learner and every classroom is unique and different. Therefore, it is difficult for only one WCF technique to match all the different people, settings and learning stages. The most efficient type of WCF depends on "a complex and dynamic interaction of linguistic and affective factors" (p. 329).

14.2 Delayed self-correction or noticing the hole enhanced

In this section, I would draw the reader's attention to the advantages found in promoting noticing opportunities while writing. My discussion of noticing opportunities in this section considers learners' reports on the NSs. Special attention is given to the findings on NS-2 from SELF-G, as they reveal the benefits of delayed self-correction.

Learners' written reports may suggest that encouraging noticing opportunities during the writing process helped learners to identify their language inaccuracies and become aware of the features they reported. This observation is made from the amount, type, way and depth of information reported on the NSs. Depth of noticing was not an aim of this study; however, in the analysis of NS-2, I realized that different task instructions in different ECCs seem to have elicited different levels of language processing. Self-provided or self-initiated feedback via self-correction demanded more language processing from SELF-G. In the DIR and REF groups, not only was correction explicitly provided but identifying their corrections in feedback was optional. More language processing in SELF-G, however, did not mean more understanding. Learners' self-correction ranged from accurate to 'inaccurate, partially

accurate, unnecessary corrections, inaccurate identification of errors causing new errors and even unnoticed errors'. What was, however, valuable for learners' language processing in the self-correction group was their hypothesis testing, calls for help and questions or comments that led them to 'notice the hole' between L1 and L2. The amount and depth of language processing arising on NS-2 in SELF-G, reinforced by learners' positive opinions about it in the exit questionnaire, make me recommend a new stage in the writing process. Different from the already known self-correction or self-revision stage (normally performed immediately after the composing stage), I would call this a 'delayed self-correction' stage, as it implies a time lapse (and even displacement) between the composing and feedback stages. Learners claimed (on NSs and questionnaires) that such a time lapse was crucial for noticing. For example: "I read my essay many times before handing it in. Surprisingly today, I realized there were words with missing letters or wrongly spelled" (Student 4-SELF). "I was able to say things in a better way today; ideas came better structured" (Student 4-SELF).

Learners' reports in SELF-G, in my view, revealed a more refined version of Nassaji's (2010) 'noticing the hole'. In other words, learners 'noticing the hole', I aimed at promoting via NS-1 in the composing stage, was better stimulated via NS-2 in the feedback stage. The large amount of language processing in SELF-G might have prepared learners to better process consecutive feedback (Swain, 1993, 1995). My proposal is explained visually below.

Current writing process:

Composing	Self-editing or immediate self-correction	Feedback or noticing the gap	Revision

Recommended stage to be incorporated in the writing process:

Composing	Self-editing or immediate self-correction	*Delayed self-correction or noticing the hole enhanced*	Feedback or noticing the gap	Revision

Delayed self-correction or noticing the hole enhanced: is an opportunity for learners' self-initiated noticing from their delayed self-correction or delayed self-provided feedback. This stage does not imply successful self-correction; however, it will allow learners to self-correct their mistakes (inaccuracies due to processing inefficiency, i.e. working memory's failure to cope with the demands of the task) and concentrate on their errors (inaccuracies due to lack of declarative knowledge, by not knowing the rule which governs a wrong item). This stage might also prepare learners to make the most out their explicit feedback to come.

Feedback or noticing the gap: is an opportunity for learners' self-initiated noticing from teacher (or other) provided feedback (external explicit feedback provided by others).

The recommended writing stage "Delayed self-correction or noticing the hole enhanced" finds support in Conti's (2004) doctoral dissertation. Conti maintains that traditional EC is teacher-centred because it fails to differentiate errors from mistakes. This leads to treating them equally. Mistakes are language inaccuracies due to processing inefficiency (working memory's failure to cope with the demands of the task). As inaccuracies due to performance they call for strategic treatment: plenty of practice and constant feedback, as stated in skill acquisition theory (McLaughlin, 1987; DeKeyser, 2007). Errors are language inaccuracies due to a lack of declarative

knowledge (not knowing the rule which governs a wrong item)**.** As inaccuracies due to lack of knowledge, they call for cognitive restructuring. Integrating noticing opportunities in the writing process as in the proposed 'delayed self-correction' may let teachers and researchers differentiate errors from mistakes in learners' interlanguage.

Continuing with noticing reports, these also displayed learners' surprise at their own discoveries. The DIR and REF groups, for instance, expressed amazement at the higher number of non-grammatical over grammatical errors they had made. Contrary to their expectations, they discovered accuracy was not only about their feared grammar rules but that non-grammatical features such as accurate vocabulary, use of connectors, attention to formal language and punctuation were equally important. Learners seemed to have gained awareness of the fact their successful writing problems lay not only in incorrect language use but also in non-native-like writing. For example:

"Most of my corrections were in writing style, making my essay more formal and native-like" (Student 8-DIR).

"My essay was correct but it was not the way a native speaker would write it" (Student 11-DIR).

"After reading the reformulated essay, I realized that it keeps the same idea as the first essay but the reformulation is more natural and fluent" (Student 6-REF).

"What attracted my attention was how different the word order between Spanish and English is" (Student 2-DIR)

"My writing is weird; I write with Spanish grammar" (Student 3-DIR).

Learners' noticing reports of their constant L1 translation reveals they also gained in awareness of this flaw.

"Everything I write is a literal translation from Spanish" (Student 13-DIR).

"I have to think as an English speaker to structure my ideas better. At present, I translate what I write" (Student 4-REF).

"My use of continuous tense was replaced by simple tense in the reformulation. I was translating" (Student 7-REF).

"I had problems in writing because I was thinking in Spanish" (Student 10-SELF).

SELF-G, on the other hand, expressed surprise at their ability to self-correct and at the large number of mistakes not noticed in their first essay. The latter was actually reported by all experimental groups. Learners explained that despite having revised their essays and written their best final version of essay 1, there were errors they were only able to perceive in the second session. For example:

"When I self-corrected [my essay] I realized there was no order in the ideas, periods and commas [were] missing. I do have to organize my ideas better" (Student 9-SELF).

"I am stunned at the number of mistakes I made, especially in spelling. I have to concentrate more when I write" (Student 9-SELF).

"I totally forgot about words I could have used" (Student 10-DIR).

"A good percentage of my mistakes were in spelling and basic things due to distraction" (Student 16-DIR).

"I wrote words the way I pronounce them; I was not careful with spelling" (Student 3-SELF).

"I had to keep in mind I was using plural, 'We' not 'I'" (Student 10-SELF).

"I do not pay enough attention to 'subjects' when I write. I know the rule for the third person singular but I do not use it" (Student 14-REF).

"I omitted letters in different words but when I was writing I did not realize about that" (Student 15-REF).

14.3 Task- and learner-related variables

From the literature review, Manchón (2011b, p. 75) concluded that *what* learners pay attention to while processing WCF depends on task-related (time on task, stage of the writing process, form-oriented vs. meaning-oriented tasks and, learners' allocation of time on task among various constituents of the composing process) and learner-related (L2 proficiency, affective and motivational individual differences) variables. Here I want to draw attention to the specific variables that emerged in this study. Concerning task-related variables, I see "allocation of time on task" and "task stage" as the most important. The non-grammatical accuracy improvement found in SELF and CONTROL groups suggest that as their noticing capacity is limited, learners need to allocate their time and attention to different language features in different task stages. Learners' attention in this study concentrated first on conveying meaning (non-grammatical features) in the composing stage. Another stage might be explored to find evidence for learners' attention to form. Swain and Lapkin (1995), for instance, found that their participants only paid attention to grammar in the editing phase.

Other task-related variables that might have influenced learners' processing of feedback in this study could include: (a) interesting topics or motivating reasons to write; (b) working on learners' own content is a task feature that emerged in the REF group; (c) rich and novel input for learners' own ideas/ content was a task feature that made reformulation successful; (d) incorporation of noticing opportunities in instruction seems to be something learners enjoyed. If noticing sheets are used for this purpose, including minimum instructions in their design might be important. Learners'

freedom in how to organize their noticing (whether they make lists, underline, build charts or categorize the error types they make) should be their choice; (e) challenging tasks, e.g. self-correction, looking for differences, correcting their peers; (f) tasks that allow learners to see their strengths were important in reformulation; (g) allowing a time lapse before self-correction was productive.

Concerning learner-related variables, examples of learners' individual differences found on noticing sheets were presented above in the section with the same title. Other learner-related variables that might have influenced learners' processing of feedback in this study could include learners' (a) engagement in communicating their content; (b) metalinguistic knowledge; (c) L2 proficiency level; (d) matching between the used WCF technique and both learners' L2 proficiency and writing stage; an intermediate learner might feel frustrated with reformulation and be unable to see the richness of this technique. For less advanced levels, DIR error correction might be more effective. Advanced learners, on the other hand, might be able to self-correct many of their mistakes, thus they may value input in reformulation.

A comment should be made about the observed learners' insufficient metalinguistic knowledge to talk about or identify their errors on the NSs. Due to lack of metalinguistic knowledge, learners frequently categorized most types of errors as lexical. In DIR-G, for instance, three students argued they had made lexical errors, yet their examples were more grammatical: good vs. well (Student 2-DIR), must vs. should (Student 13-DIR), if vs. whether (Student 14-DIR). Lack of metalanguage also led some learners to identify grammatical structures inaccurately, and consequently fail to identify their needs. For example:

"I need to review when to use the preposition 'the'" (Student 10-REF)

"I have problems with passive voice" (Student 12-DIR). The learner actually referred to a subject + verb agreement error in a conditional sentence.

Researchers studying learners' writing while thinking aloud (e.g. Armengol & Cots, 2009; Van Weijen, Van den Bergh, Rijlaarsdam & Sanders, 2009) have found evidence for learners' attention to form though not referring to mentioning explicit rules. Researchers have attributed this to limitations of the think-aloud technique. Analyses of the NSs in this study also found evidence of learners' attention to form. Contrary to the above-mentioned studies, participants in this study did mention explicit rules. However, their written reports reflected inaccurate rules due to their lack of metalinguistic knowledge. Absence of metalinguistic knowledge then might also explain the no mentioning of explicit rules in the above referred studies.

Chapter 15: Implications of the Study

15.1 Theoretical implications

15.1.1 Truscott's claims and experimental research

For more than two decades, researchers have tried to refute Truscott's claim for the inefficiency of WCF for grammar improvement. Despite numerous attempts, the role WCF plays in grammar learning is still unresolved. Problems with early studies trying to negate Truscott's claim included methodological and ethical flaws that recent studies have overcome. Researchers are, however, still unable to reject Truscott's claim. I see three reasons for this. First, the type of feedback and the context of WCF that Truscott refers to have been taken for granted. Truscott's claims refer to real classroom contexts whose variables are numerous and difficult to approach in experimental research. And researchers have chosen the type of WCF and grammar structure(s) that are convenient for their analysis. This has resulted in copious investigations, on different structures, testing different WCF techniques, within different contexts and with different writing tasks. Such variations have caused research results to be fragmented, not generalizable and limited. Second, this situation derives from the requirements of experimental research: (a) controlled environments imply isolating the writing process from other classroom variables, such as multiple drafting, whole-class feedback and instruction; (b) controlled variables have led to a focus on specific grammar targets, specific language proficiency levels and a reduced number of words; (c) controlled tasks have also translated into unauthentic controlled writing tasks (picture descriptions and picture sequence narratives where writing is approached as a means to practise grammar rather than an end itself); (d) controlled procedures have also translated into pre-, post- and delayed post-test sequences. The demands of experimental research have, to a certain extent, reduced the ecological

validity of research designs, making them distant from classroom realities. Third, published WCF research seems to be inclined towards studies supporting WCF (Truscott, 2007). Ellis, et al.´s (2008) study, for instance, is presented as dealing with comprehensive error correction, despite the authors themselves making it clear they used semi-comprehensive EC. Van Beuningen's (2011) support for the efficacy of comprehensive EC in accuracy improvement is also generalised, despite the uncommon naturalistic teaching context of the study. In Truscott's (1996) meta-analysis of WCF studies, he stated published work had been unfair in favouring the supportive effects of correction and offering little space for different views. Bias in favour of correction research makes those studies "look better than they actually are" (Truscott, 2007, p. 267).

15.1.2 Support for Truscott's claim

The lack of statistically significant results for grammatical features in this study supports Truscott's claim for the inefficiency of comprehensive WCF for grammatical accuracy improvement. Different from Truscott, however, I do not claim that WCF is inefficient. To me, the results show rather that although non-grammatical features emerged (a) as the most reported noticed features in the composing and feedback stages, and (b) as the features most amenable to correction and accuracy improvement, they have received little attention. The results do not suggest teachers should stop providing or investigating WCF, they suggest what Evans et al. (2010, p. 446) remarked on, that we do not want to know whether providing WCF is efficient or not, we want to know how we can best help students write more accurately. If non-grammatical errors are more noticeable, more amenable to correction and contribute to writing accuracy, teachers could pass this responsibility over to learners. Another reason why I do not claim that WCF is inefficient concerns my point about experimental studies looking for

learners' attention to grammar in the wrong writing stage (the second draft in the composing stage rather than in the editing stage).

15.1.3 Considerations of learning-to-write and writing-to-learn

L2 writing is normally considered an ability to acquire, teach and assess. Contrary to this assumption, Cumming (2011) states that "L2 writing is also a means, context and basis for learning, both of language and of writing" (p. *x*). With this statement Cumming draws attention to the importance of exchanging ideas between theory, research and teaching practice in SLA and written composition. This study has attempted to promote this exchange. I framed my research within a writing-to-learn, language for acquisition and cognitive interactionist framework. I targeted a foreign language context and struggled to create a pedagogically acceptable design within the limits of experimental research. My design confronted me with teachers' and learners' conceptions, practices and expectations about what teaching, learning, writing and language are. If something became clear in this study it is that "there is … much more to writing competence than grammatical accuracy" (Lee, 2011 p. 386). L2 learners' writing difficulties (in the writing-to-learn language perspective) reach the learner-to-write dimension simply because appropriate text construction, as syntax, semantics, cohesion and coherence are all text components. Good writing is in fact an amalgam of cognitive skills and multiple-language written literacy capacities. In the WLL, L2 writing demands processing at two levels: language accuracy (writing-to-learn language dimension) and clear presentation and organization of ideas (learning-to-write). Both dimensions are necessary and complement each other.

The LW, WLC and WLL dimensions introduced in the literature review are decisive to understand different conceptions, uses, classroom contexts and goals of writing. Research on writing from each of these perspectives is important. However,

"more often than not, the three views are closely related" (Ortega, 2011, p. 244). Further looking for connections between different writing spheres leads me to what Ortega (2012) refers to as 'the role of L2 proficiency in L2 writing', an intersection point between SLA and L2 writing that a group of scholars have started to study. The scope of this thesis does not allow me to go deeper into this topic. However, Schoonen, Snellings, Stevenson and Van Gelderen (2009) state that "the relationship between L1 and FL writing proficiency is without doubt mediated by FL linguistic knowledge … the issue of how and to what extent these three constructs interact is still not settled" (p. 82).

15.2 Pedagogical implications

15.2.1 Effects of different ECCs

Findings about how ECCs influence the amount of noticing and the type of features noticed is central for teachers. As many techniques as possible should be tried to help learners understand what writing is about and what techniques are available to reach their goals. The straightforwardness and comprehensiveness of DIR error correction make it effective for beginner learners who do not have enough L2 knowledge to identify their errors by themselves. However, when teachers aim to raise awareness of non-grammatical features, reformulation seems to be more efficient. In the end, no best WCF technique exists, learners' individual differences and the multi-stage nature of writing will determine which WCF technique is best to use in a given context.

15.2.2 Reassurance of delayed self-correction

Delayed self-correction emerged as a stage to be further stimulated. Its effectiveness in noticing was surprising and satisfactory for both researcher and

learners. Despite the use of checklists or allowing time for self-correction in writing classes, its value had not been thought as promising as it proved in this study. Different from the already practised self-correction or self-revision, a time lapse was found important to increase the potential of self-correction. The positive effects of this time lapse for learners' focus on form became evident. Incorporating delayed self-correction to the writing process may make the correction process more efficient by allowing learners to self-correct their mistakes (inaccuracies due to processing inefficiency) and teachers to focus on learners' errors (inaccuracies due to lack of declarative knowledge). Delayed self-correction might also prepare learners to make the most out of their explicit external feedback to come. Needless to say, learners will be able to exploit their own explicit knowledge and teachers may make their correction task less demanding. Written reports of features learners noticed do not mean those features were learned. Written reports are only evidence they were noticed and constitute merely the first stage towards awareness. The large amount of accurate self-correction reported on NS-2, especially in SELF-G, reveals learners can still improve much by themselves when given the chance. Polio (2012) suggests that future research should consider "how and when WCF can be made more effective" (p. 386). Delayed self-correction is suggested.

15.2.3 From language teachers to teachers of writing

Giraldo de Lodoño and Perry (2008) stress that teachers' role in FL writing is to help learners become independent writers. This goal requires first that FL teachers change their own assumption as 'language teachers' to a new one, 'teachers of writing'. The latter implies engaging writers in a constructive process with multiple stages: drafting, composing, rewriting, self-revision, peer correction, content revision, form revision, proofreading, editing. Mantello (1997) suggests that multiple stages of

writing require teachers to play different roles depending on their correction purposes at different stages in the writing process.

15.2.4 Writing: skill vs. knowledge

The acknowledgement of writing as a skill (McLaughlin, 1987; DeKeyser, 2007) might also allow teachers to approach writing like any other ability when delivering error correction. The skill acquisition theory introduced in the literature review suggests that in sports, the arts (music, ballet) or everyday life skills (e.g. driving, cooking, playing an instrument), learners work on a few skills at a time, correction is not general but specific. In the development of skills, explicit knowledge requires practice and feedback. Dose-based, systematically selective marking seems to be a good error correction practice (Lee, 2004, p. 301). Priorities must be established to know what to correct and in which sequence. Zamel (1985), recommends meaning level be addressed first. Results of this study, and those of Swain and Lapkin (1995) hint this might be a more fruitful pedagogical option.

15.2.5 Potential of reformulation in FL contexts

Learners' motivation in reformulation was revealed in the enthusiasm they used to talk about their noticing. Long (2007) explains that because learners' correct version of their own sentences requires less attention to meaning, more attention is given to form. Reading their own content was engaging for learners. They were proud of their ability to communicate and were interested in paying attention to features to improve their writing beyond form: referencing devices, lexis, synonyms, variety of connectors, formal language, punctuation. Giraldo de Lodoño and Perry (2008) state that writing in FL classrooms is 'so handicapped that … students tend to regard details such as punctuation and use of capital letters as unimportant' (p. 120). Learners'

engagement with processing via reformulation may then justify further research into this technique and more frequent implementation.

Studies have found that the effects of WCF on learning are also determined by learners' engagement with processing (Bitchener, 2008; Bitchener & Knoch, 2008). Regarding this, NS-2 and answers to Q1 in the questionnaire demonstrated that learners considered REF to be the most engaging ECC. If REF was so efficient for learner-initiated noticing of overall features and promoted much introspection, why then did this efficiency not translate into accuracy? Learners' engagement produced by reformulation may well have a delayed effect on learning that non-longitudinal designs are unable to show. The incorporation of text construction features might not be as immediate as incorporation of lexis or grammar structures. Schmidt's (2001) also remarks that noticing does not mean learning. Noticing is only the first in a series of cognitive processes in the input-to-intake transformation.

Helping learners become independent writers entails making learners responsible for their "self-monitoring and self-correction processes" (Porte, 1993, p. 43). If having learners rewrite their revised text is the only way teachers know to draw learners' attention and it has not worked, then Polio (2012) suggests teachers have to explore other methods to encourage noticing opportunities. The results of this study have shown that encouraging noticing opportunities, delayed self-correction and reformulation are efficient techniques for promoting autonomy and learner-initiated noticing. The incorporation of these features into instruction may guarantee that the first step in the learning process is taken and that learning becomes more active and learner-centred.

Chapter 16: Limitations of the Study

There were two main challenges in this research and various limitations. First, the operationalization and measurement of noticing, particularly the 'absence of noticing' and the difficulty to tackle comprehensive ECCs, were the main challenges. Weaknesses in the design, such as operationalization of the absence of noticing as blank NSs, confirmed Leow's (2000) claim about the difficulty of measuring the non-appearance of noticing. Blank NSs did not mean lack of noticing, simply that learners did not report their noticing. Previous studies with similar operationalization of this construct came to similar conclusions (Mackey, 2006). Some evidence of noticing without reporting was actually found. Student 7 in REF-G, for instance, faced specific problems while composing. However, he did not report it on his NS-1. It was only when I read his comment on NS-2 that I returned to his NS-1 and realised it was blank. The student's comment on NS-2 was "In the sentence … I struggled much to express my idea. When I saw my reformulated essay I realized I was not even close!"

An important limitation of the study was the lack of at least one additional ECC that received feedback exclusively and had no noticing opportunities. The absence of this ECC does not allow for strong inferences about the value of noticing. As a quasi-experimental study within certain time constraints, this study does not overcome the limitations of current research. It is not a longitudinal study; feedback is not sustained; accuracy measurement by error ratio is not the best way to determine writing development, an instrument to determine participants' English proficiency more accurately before the study would also be pertinent. Besides the above limitations, the use of noticing sheets to obtain introspective cognitive data has been criticized for superficially connecting noticing to the facts that prompted them. Input processing is said to occur momentarily, whereas the completion of noticing sheets

takes longer. Thus, there is the possibility of forgetting an experience at the time of reporting. Observations of learners' behaviours in this study, however, make me question this criticism. Input processing does not necessarily occur momentarily. Leow (1997) and Robinson et al. (2011) explain that input processing is rather a sequence of momentary awareness level episodes. Might this explain the success of delayed SELF-correction?

Regarding post-exposure questions, some researchers (Mackey, 2006; Robinson, 1995) claim their usefulness depends on learners' capacity to separate what they notice during learner-stimuli interaction from what they notice while completing a questionnaire. Most researchers consider a mixture of online and offline procedures as the best option to measure cognitive processes. However, even in Mackey's 2006 study, which included a variety of instruments, the combination of procedures created new limitations.

Lack of inter- and intra-reliability made comprehensive error correction an unfeasible task. A whole section in this thesis was devoted to that methodological problem. Studies with comprehensive error correction are indeed necessary in SLA. Ellis et al. (2008), Van Beuningen (2011) and Polio (2012), among others, have called for this type of research. Provided comprehensive EC is possible, Truscott (1996) still warns researchers that this might not be a solution (p. 353), experimental research requires control of many other variables.

Finally, the need to digitize my data to avoid onerous pen-and-paper work added a computer-mediated writing variable to this study. This might have had effects on learners' performance. Excluding studies where computer-mediated communication is their aim, most existing research in WCF is performed with pen and paper. If research in the latter type of studies has been computer-mediated, little or no

reference has been made to the way writing on a computer may alter the way learners write, teachers correct and people learn in general. Neither computer-mediated nor pen-and-paper writing is better; they are people's preferences. However, students' preferring pen-and-paper writing might have been a disadvantage for those preferring computer-mediated writing in this study. Despite their being the digital generation, some students found it challenging to write on a computer. One student wrote on his noticing sheet-2 that some of his/her mistakes were typing rather than spelling mistakes; a second student said it had been puzzling for her/him to write English on a computer as s/he claimed errors were easier to perceive on paper than on a computer monitor. S/he also argued that writing in English on a PC was not common in EFL contexts where L2 writing, even on paper, is scarcely practised. Another student (DIR-G) mentioned s/he was not used to the way corrections look in a Word document. S/he argued s/he had first to get used to the format. Concerning teacher-raters, differences and preferences also emerged. Only one out of three participant teachers felt comfortable correcting on a computer. The others said they needed to print the essay, correct it on paper and only then categorize the errors. Teachers claimed they were used to correcting in that way because their full-time job obliged them to correct learners' writing anywhere, anytime, without computer support. As for the researcher, I was in charge of correcting learners' essays in the DIR group. The computer-mediated correction made clear, tidy corrections possible and allowed for longer and more accurate amendments, something unmanageable in pen-and-paper corrections. While correcting, I considered this an advantage; however, I was unaware of the effects computer-mediated correction might have on learners. To me, the nature of the writing task itself (uncontrolled opinion essays demanding corrections beyond accuracy at a sentence level), and the way a computer facilitated correction, might

have unintentionally modified traditional pen-and-paper direct error correction. Future studies should then consider how computer-mediated writing, correction and learning might in general be influenced by this component.

Chapter 17: Conclusion

This research aimed to design a pedagogically acceptable study on the effects of different ECCs on learner-initiated noticing. With this in mind, I made an effort to tackle comprehensive error correction in an open, uncontrolled, learner-centred, syllabus-based writing task. Comprehensive EC was eventually replaced by semi-comprehensive EC. However, the challenges of comprehensive EC for experimental research have been reported, hoping to set better guidelines for future studies. The nature of learner-initiated noticing (with no predetermined target), the characteristics of uncontrolled open writing tasks, and the complexity of comprehensive EC made my research full of challenges and discoveries. Regarding 'corrective feedback', Ortega (2012) sees pace and permanence as advantages of writing that need to be further exploited. In this design, I made the most of pace by setting no time restrictions other than a two-hour session for task performance. No time restrictions allowed learners to plan, rewrite, recall their explicit knowledge and revise their texts. The inclusion of noticing sheets in the composing and feedback stages also maximized the slow pace of writing. Delayed self-correction exploited the advantages of permanence too. By offering a time lapse between the composing and self-correction sessions, the design increased the possibilities for learners to notice gaps and pick up their writing with clearer minds, factors that may facilitate consolidation in the future.

The findings were more than originally planned. The study showed that working with open, uncontrolled, learner-centred tasks like opinion essays is possible. Results from an exploration of this type of task may support Manchón's (2011b) observation regarding the features of language that learners pay attention to while receiving feedback. She states that research has shown that "the more open the task the more focus on lexis and less attention paid to grammar" (p. 72). My results agree

with this conclusion. However, as spelling was included, it came up as the feature most noticed and amenable to correction.

The results also highlight the value of noticing and delayed self-correction to help learners develop autonomy and become independent writers. The study suggests that noticing opportunities may be as important as error correction conditions. And that a noticing stage in the writing process might be beneficial. For LLPW to help consolidation, the writer's control of L2 knowledge and the generation of new L2, two things might be necessary. First, experimental studies should try to operationalise writing as a process. To me, it seems that current experimental designs deal with writing as a product. The composing and rewriting stages in most designs actually correspond to two drafting moments in the composing stage. Second, a noticing stage could be formally added to the writing process. The various cognitive processes that writing entails and the limited nature of attention oblige learners to distribute their attention across different writing stages. If research has shown that non-grammatical features are noticed first and attention to form occurs in a subsequent stage then, Truscott might be partially right, in that grammar correction of writing in the composing stage is futile because in this stage, learners are struggling with the formulation of meaning. The question is, will grammar correction be ineffective in the editing stage as well?

Regarding ECCs, ***delayed self-correction*** emerged as the type of WCF that elicited the most attention to form; ***direct error correction*** arose as the one leading to more retention, and ***reformulation*** was the most engaging one. My study supports Truscott's claim for the inefficiency of (semi-comprehensive in this case) WCF for grammatical accuracy. Different from Truscott, however, I do not sustain that WCF is inefficient. In my view, traditional forms of WCF may be improved concerning time

and energy consumption. Lee (2011) points out that teachers need to work smarter and not harder in responding to student writing. Integrating my proposed delayed self-correction into the writing process may help to achieve this goal. Conti (2004) suggests learners should be trained in the development of noticing strategies if they are to become independent writers. Adding this stage to the writing process may contribute to Conti's suggestion. Learning does not take place by simply looking at teachers' correction. Due to learners' limited processing capacity, they need to be trained in attention and noticing to "be selective and … strategically allocated and managed" (Izumi, 2013 p. 35). The findings also suggest that when learners are given the freedom to choose what to notice, attention to form is not a priority. Thus I draw attention to the small value attached to non-grammatical features in WCF research. The efficacy of error correction for spelling accuracy improvement should in my view, be enough to justify the value of WCF.

Schmidt (1990) says noticing is only the first stage in the more complex learning process. Proponents (Anderson, 2000; Johnson, 1988) of skill theory suggest noticed language features will need extensive practice and constant feedback to be acquired. Longitudinal studies are essential to find out what the long-term effects are. However, if noticing comes from the learner rather than being externally imposed, it might be more effective.

When I started this research, I affirmed that learners are the ones who decide what, when, how much and how deeply they notice. Awareness of the endless *internal* (affective: motivation, aptitude, beliefs; cognitive: developmental readiness, limited processing capacity, L1, current L2 knowledge), *external* (instruction, feedback, interaction and task demands) and *input* (frequency, perceptual saliency, communicative value of form) factors influencing the noticing process (Izumi, 2013;

Leow, 2013) made me reconsider my view and further explore Izumi's conclusion "Learners are not free to notice anything and everything they wish to notice … noticing is not always voluntarily controlled, either. [Noticing] is constrained and regulated by many properties in the input, as well as many factors internal to the learners" (p. 35).

Directions for future research could include (a) a more systematical approach to comprehensive EC. If comprehensive error correction is feasible for research, agreement might first be necessary on the categorisation of different error types. (b) The interaction between error categories and type of WCF needs more attention too. As Bitchener (2012) states, "the extent to which written CF can effectively target different types of error is in the very early stages of investigation" (p. 356). (c) Longitudinal studies on learners' accuracy performance after receiving noticing treatment are essential. Learners' testimonials and their affluent written reports in NSs showed that delayed self-correction might have improved the conditions to receive subsequent feedback. Future studies could explore this initial finding further. (d) Future research should also consider deeper enquiry into the relationship between noticing and individual learner differences. Schmidt (2010) and Godfroid (2010) have drawn attention to individual learner differences determining what we notice, what we learn. Gardner (1988) suggests motivation allows learners to reach higher levels of awareness. Bley-Vroman (1989), Krashen, (1981) and Reber (1993) have also proposed different relationships among aptitude, noticing and SLA. (e) as the most noticed feature, lexis should be further enquired too.

Like Evans et al. (2010) and Lee (2011), I maintain that teachers should make EC an essential component of teaching and learning and should continue to enquire into the best ways to help learners improve their writing skill. Evans et al. (2010) state

"there are scientific and ethical reasons … to continue research on correction" (447). After all, Johnson (1988) states that "the question of how to provide successful feedback is no less perplexing than the question of how to facilitate successful ... learning" (p. 95).

References

Adams, R. (2003). L2 output, reformulation and noticing: Implications for IL development. *Language Teaching Research, 7*(3), 347–376.

Allwright, R.L., Woodley, M.P., & Allwright, J.M. (1988). Investigating reformulation as a practical strategy for the teaching of academic writing. *Applied Linguistics, 9*(3), 236–256.

Anderson, J. (1985). *Cognitive psychology and its implications* (2nd ed.). New York, NY: Freeman.

Armengol, L., & Cots, J. (2009). Attention processes observed in think-aloud protocols: Two multilingual informants writing in two languages. *Language Awareness, 18,* 259–276.

Ashwell, T. (2000). Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing, 9*, 227–258.

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*, 102–118.

Bitchener, J. (2012). A reflection on 'the language learning potential' of written CF. *Journal of Second Language Writing, 21*, 348–363.

Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research, 12*, 409–431.

Bitchener, J., & Knoch, U. (2009a). The contribution of written corrective feedback to language development: A ten-month investigation. *Applied Linguistics, 31*, 193–214.

Bitchener, J., & Knoch, U. (2009b). The relative effectiveness of different types of direct written corrective feedback. *System 37*, 322–29.

Bitchener, J., & Knoch, U. (2010). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing, 19*(4), 207–217.

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing, 14*, 191–205.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language*

acquisition (pp. 41–68). Cambridge: Cambridge University Press.

Bruton, A. (2009). Improving accuracy is not the only reason for writing, and even if it were… *System, 37,* 600–613.

Camps, D., Villalobos, J., & Shea, J. (2012). Understanding EFL students' errors: An insight towards their interlanguage. *MEXTESOL Journal, 36*(1), 2012.

Carr, T., & Curran, T. (1994). Cognitive factors in learning about structured sequences: Applications to syntax. *Studies in Second Language Acquisition, 16,* 205–230.

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing, 12,* 267–296.

Cohen, A.D. (1983). *Reformulating second-language compositions: A potential source of input for the learner*. (ERIC ED 228 866).

Conti, G. (2004). *Metacognitive enhancement and error correction: An investigation in the impact of self-monitoring strategies on L2 Italian student writing* (Unpublished doctoral thesis). University of Reading, United Kingdom.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioural and Brain Sciences, 24,* 87–114.

Cumming, A. (1983). *Teachers' procedures for responding to the writing of students of English as a second language*. Paper presented at the 16th Annual Canadian Council of Teachers of English Convention, Montreal.

Cumming, A. (1990). Metalinguistic and ideational thinking in second language composing. *Written Communication, 7*(4), 482–511.

Cumming, A. (2011). Preface. In R.M. Manchón (Ed.), *Learning-to-write and writing to-learn in an additional language* (pp. VIII–XII). Amsterdam: John Benjamins Publishing Company.

DeKeyser, R. (Ed.). (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge: Cambridge University Press.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R., Basturkmen, H., & Loewen, S. (2001). Pre-emptive focus on form in the ESL classroom. *TESOL Quarterly, 35,* 407–32.

Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System, 36,* 353–371.

Evans, N.W., Hartshorn, K.J., McCollum, R. M., & Wolfersberger, M. (2010). Contextualizing corrective feedback in second language writing pedagogy. *Language Teaching Research, 14*(4), 445–463.

Fathman, A., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 178–190). Cambridge: Cambridge University Press.

Fazio, L. L. (2001). The effect of corrections and commentaries on the journal writing accuracy of minority- and majority-language students. *Journal of Second Language Writing, 10,* 235–249.

Ferris, D. (2004). The 'grammar correction' debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime... ?). *Journal of Second Language Writing, 13*, 49–62.

Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81–104). Cambridge: Cambridge University Press.

Ferris, D. (2010). Second language writing research and written corrective feedback in SLA. *Studies in Second Language Acquisition, 32*, 181–201.

Ferris, D. (2011). *Treatment of error in second language student writing.* USA: University of Michigan Press.

Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes. How explicit does it need to be? *Journal of Second Language Writing*, *10*, 161–184.

Frantzen, D. (1995). The effects of grammar supplementation on written accuracy in an intermediate Spanish content course. *Modern Language Journal, 79*, 329–344.

Gardner, R.C. (1988). The socio-educational model of second-language learning: Assumptions, findings, and issues. *Language Learning, 38*, 101–126.

Gass, S. M. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics, 9*, 198–217.

Gass, S.M., & Selinker, L. (2008). *Second Language Acquisition. An introductory course*. (3rd ed.). New York, NY: Routledge.

Giraldo de Londoño, M.A., & Perry, R.A. (2008). Treating ESL writing errors:

Balancing form and content. *HOW, A Colombian Journal for Teachers of English 15,* 107–123.

Godfroid, A. (2010). *Cognitive processes in second language acquisition: The role of noticing, attention and awareness in processing words in written L2 input* (Unpublished doctoral dissertation). Vrie Universiteit Brussel, Brussels.

Godfroid, A., Housen, A., & Boers, F. (2010). A procedure for testing the noticing hypothesis in the context of vocabulary acquisition. In M. Pütz & L. Sicola, (Eds.), *Cognitive processing in second language acquisition. Inside the learner's mind* (pp. 169–198). Amsterdam: John Benjamins Publishing Company.

Guénette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing, 16,* 40–53.

Hanaoka, O. (2007). Output, noticing, and learning: An investigation into the role of spontaneous attention to form in a four-stage writing task. *Language Teaching Research, 11*(4), 459–479.

Hanaoka, O., & Izumi, S. (2012). Noticing and uptake: Addressing pre-articulated covert problems in L2 writing. *Journal of Second Language Writing, 21,* 332–347.

Hartshorn, K. J., Evans, N. W., Merrill, P. F., Sudweeks, R. R., Strong-Krause, D., & Anderson, N. J. (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL Quarterly, 44*, 84–109.

Housen, A., & Pierrard, M. (2005). Instructed second language acquisition: Introduction. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 1–26). Berlin: Mouton de Gruyter.

Hyland, K. (2003). *Second language writing*. USA: Cambridge University Press.

Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis. An experimental study of ESL relativization. *Studies in Second Language Acquisition, 24*, 541–577.

Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics, 24*, 168–196.

Izumi, S. (2013). Noticing and L2 development: Theoretical, empirical and pedagogical issues. In J.M. Bergsleithner, A.N. Frota & J.K. Yoshioka (Eds.), *Noticing and second language acquisition: studies in honor of Richard Schmidt* (pp. 25–38). Honolulu: University of Hawaii, National Foreign Language Resource Center.

Izumi, S., & Bigelow, M. (2000). Does output promote noticing and second language

acquisition? *TESOL Quarterly, 34*, 239–278.

Johnson, K. (1988). Mistake correction. *ELT Journal, 42*(2), 89–96.

Kepner, C. (1991). An experiment in the relationship of types of written feedback to the development of second language writing skills. *Modern Language Journal,*

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing, 21,* 390–403.

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.

Krashen, S. (1983). *The natural approach. Language acquisition in the classroom*. USA: Pergamon Press.

Krashen, S. (1985). *The input hypothesis*. London: Longman.

Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing, 17*, 48–60.

Lalande, J. F. (1982). Reducing composition errors: An experiment. *Modern Language Journal, 66*, 140–149.

Lapkin, S., & Swain, M. (1990). French immersion research agenda for the 90s. *Canadian Modern Language Review. 46,* 638–369.

Lapkin, S., Swain, M., & Smith, M. (2002). Reformulation and the learning of French pronominal verbs in a Canadian French immersion context. *The Modern Language Journal, 86*(4), 485–507.

Lázaro, A. (2009). Reformulation and self-correction: Testing the validity of correction strategies in the classroom. *RESLA Revista Española de Linguística Aplicada, 22,* 189–215

Lee, I. (2004). Error correction in L2 secondary writing classrooms: The case of Hong Kong. *Journal of Second Language Writing, 13,* 285–312.

Lee, I. (2005). Error correction in L2 writing classrooms: What do students think? *TESL Canada Journal, 22*(2), 1–16.

Lee, I. (2011). Working smarter, not working harder: Revisiting teacher feedback in the L2 writing classroom. *The Canadian Modern Language Review, 67*(3), 377–399.

Lee, J. F., Cadierno, T., Glass, W. R., & Van Patten, B. (1997). The effects of lexical and grammatical cues on processing past temporal reference in second language input. *Applied Language Learning, 8,* 1–23.

Leow, R. (1997). Attention, awareness and foreign language behaviour. *Language Learning, 47*(3), 467–505.

Leow, R. (2000). A study of the role of awareness in foreign language behaviour. Aware versus unaware learners. *Studies in Second Language Acquisition, 22,* 557–584.

Leow, R. (2013). Schmidt's noticing hypothesis: More than two decades after. In J.M. Bergsleithner, S. N. Frota, & J.K. Yoshioka (Eds.), *Noticing and second language acquisition: studies in honor of Richard Schmidt* (pp. 11–23). Honolulu: University of Hawaii, National Foreign Language Resource Center.

Long, M. H. (2007). Recasts: The story thus far. In M. H. Long (Ed.), *Problems in SLA* (pp. 75–116). Mahwah, NJ: Lawrence Erlbaum.

Lynch, T., & Maclean, J. (2001). A case of exercising: effects of immediate task repetition on learners' performance. In M. Bygate, P. Skehan & M.Swain (Eds.) Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing. (pp. 141–162). New York: Routledge.

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics, 27*(3), 405–430.

Manchón, R.M. (2011a). The language learning potential of writing in foreign

language contexts: Lessons from research. In T. Cimasko & M. Reichelt (Eds.), *Foreign language instruction. Principles and practices* (pp. 44–64). USA: Parlor Press.

Manchón, R.M. (2011b). Writing-to-learn the language. In R.M. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 61–82). Amsterdam: John Benjamins Publishing Company.

Manchón, R.M. (2013). Teaching writing. In C. Chapelle (Ed.), *The encyclopaedia of applied linguistics* (pp. 2−7). USA: Blackwell Publishing.

Mantello, M. (1997). Error correction in the L2 classroom. *The Canadian Modern Language Review, 54*(1), 127–131.

Matsumoto, K. (1993). Verbal-report data and introspective methods in second language research: State of the art. *RELC Journal. A Journal of Language Teaching and Research in Southeast Asia, 24*, 32–60.

McLaughlin, B. (1987). *Theories of second language learning.* London: Edward Arnold.

McLaughlin, B., Rossman, T., & McLeod, B. (1983). Second language learning: An information-processing perspective. *Language Learning, 33*, 135–157.

Murphey, T. (1993). Why don't teachers learn what learners learn? Taking the

guesswork out with action logging. *English Teaching Forum, 31*(1). Retrieved 8 August 2016 from: https://kandaeli.academia.edu/TimMurphey

Murphy, T., & Falout, J. (2010). Critical participatory looping: Dialogic member checking with whole classes. *TESOL, Quarterly*, *44*(4), 811–821.

Nassaji, H. (2010). The occurrence and effectiveness of spontaneous focus on form in adult ESL classrooms. *The Canadian Modern Language Review, 66*(6), 907–933.

Nassaji, H., & Swain, M. (2000). A Vygotskian perspective on corrective feedback in L2: The effect of random versus negotiated help on the learning of English articles. *Language Awareness, 9*(1), 34–51.

Neumann, O. (1996). Theories of attention. In O. Neumann & A. F. Sanders (Eds.), *Handbook of perception and action, Volume Three: Attention* (pp. 389–446). London: Academic Press.

Ortega, L. (2011). Reflections on the learning-to-write and writing-to-learn dimensions of second language writing. In R.M. Manchón (Ed.), *Learning-to write and writing-to-learn in an additional language* (pp. 237–250). Amsterdam: John Benjamins Publishing Company.

Ortega, L. (2012). Epilogue: Exploring L2 writing–SLA interfaces. *Journal of Second Language Writing, 21,* 404–415.

Park, E.S. (2011). Learner-generated noticing of written L2 input: What do learners notice and why? *Language Learning, 61*(1), 146–186.

Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: John Benjamins.

Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing, 21*, 375–389.

Polio, C., Fleck, C., & Leder, N. (1998). If only I had more time: ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing, 7*(1), 43–68.

Porte, G. K. (1993). Mistakes, errors, and blank checks. *English Teaching Forum, 31*(1), 42–44.

Qi, D. S., & Lapkin, S. (2001). Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing*, *10*(4), 277–303.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York, NY: Oxford University Press.

Reinders, H. W. (2005). *The effects of different task types on L2 learners' intake and*

acquisition of two grammatical structures (Unpublished doctoral thesis). University of Auckland, Auckland.

Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, *20*, 83–95.

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning*, *45*(2), 283–331.

Robinson, P., Mackey, A., Gass, S., & Schmidt, R. (2011). Attention and awareness in second language acquisition. In S. Gass & A. Mackey (Eds.), *Handbook of second language acquisition* (pp. 247–267). London: Routledge.

Sachs, R., & Polio, C. (2007). Learners' uses of two types of written feedback on an L2 writing revision task. *Studies in Second Language* Acquisition, *29*, 67–100.

Santos, M., López-Serrano, S., & Manchón, R.M. (2010). The differential effect of two types of direct written corrective feedback on noticing and uptake: Reformulation vs. error correction. *International Journal of English Studies, 10*(1), 131–154.

Shapiro, K. L., Arnell, K. M., & Raymond, J. E. (1997). The attentional blink. *Trends in Cognitive Sciences*, *1*, 291–296.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied*

Linguistics, 11, 129–158.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics, 13,* 206–226.

Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165–209). London: Academic Press.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu, Hawai'i: University of Hawai'i, Second Language Teaching and Curriculum Centre.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010* (pp. 721–737). Singapore: National University of Singapore, Centre for Language Studies.

Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in second language: A case study of an adult learner of Portuguese. In R. R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237–326).

Rowley, MA: Newbury House.

Schoonen, R., Snellings, P., Stevenson, M., & Van Gelderen, A. (2009). Towards a blueprint of the foreign language writer: The linguistic and cognitive demands of foreign language writing. In R. M. Manchón (Ed.), *Writing in foreign language contexts learning, teaching, and research* (pp. 77–129). UK: Multilingual Matters.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics, 10,* 219–231.

Semke, H. D. (1984). Effects of the red pen. *Foreign Language Annals, 17,* 195–202.

Sharwood Smith, M. (1981). Consciousness raising and second language learner. *Applied Linguistics, 2,* 159–169.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition, 15,* 165–179.

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly, 41*, 255–283.

Sheen, Y. (2010). Differential effects of oral and written corrective feedback in the ESL classroom. *Studies in Second Language Acquisition, 32,* 203–234.

Sheen, Y., Wright, D., & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System, 37,* 556–569.

Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal, 23,* 103–110.

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford: Oxford University Press.

Skehan, P. (2003). Task-based instruction, *Language Teaching, 36,* 1–14.

Storch, N. (2009). *Comparing type of feedback and processing mode: Pair versus individual processing of feedback on writing*. Paper presented at the Second Language Writing Symposium, Phoenix, Arizona.

Storch, N. (2010). Critical feedback on written corrective feedback. *International Journal of English Studies, 10*(2), 29–46.

Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing. Case Studies. *Studies in Second Language Acquisition, 32,* 303–334.

Suzuki, W. (2008). *The effect of written languaging combined with feedback on*

second language writing. Paper presented at the AAAL, Annual Conference, Washington, DC.

Swain, M. (1985). Communicative competence: some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Rowley, MA: Newbury House.

Swain, M. (1993). The output hypothesis: just speaking and writing aren't enough. *Canadian Modern Language Review, 50,* 158–164

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: studies in honour of H. G. Widdowson* (pp. 125–144). Oxford: Oxford University Press.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step toward second language learning. *Applied Linguistics*, *16,* 371–391.

Swain, M., & Lapkin, S. (2002). Talking it through: Two French immersion learners´ response to reformulation. *International Journal of Educational Research*, *37,* 85–304.

Swain, M., & Lapkin, S. (2007). The distributed nature of second language learning:

Neil's perspective. In S. Fotos & H. Nassaji (Eds.), *Form-focused instruction and teacher education. Studies in honour of Rod Ellis* (pp. 73–86). Oxford: Oxford University Press.

Tocalli-Beller, A., & Swain, M. (2005). Reformulation: The cognitive conflict and L2 learning it generates. *International Journal of Applied Linguistics, 15*(1), 5–28.

Tomlin, R., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 16,* 183–203.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46*(2), 327–369.

Truscott, J. (1998). Noticing in second language acquisition: A critical review. *Second Language Research, 14*(2), 103–135.

Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing, 16,* 255–272.

Truscott, J., & Hsu, A. Y. (2008). Error correction, revision, and learning. *Journal of Second Language Writing, 17,* 292–305.

Truscott, J., & Sharwood Smith, M. (2011). Input, intake, and consciousness: the quest for a theoretical foundation. *Studies in Second Language Acquisition, 33*(4), 497–528.

Van Beuningen, C. (2011). *The effectiveness of comprehensive corrective feedback in second language writing* (Unpublished PhD thesis). University of Amsterdam, Amsterdam.

Van Beuningen, C., De Jong, N., & Kuiken, F. (2008). The effect of direct and indirect corrective feedback on L2 learners' written accuracy. *ITL International Journal of Applied Linguistics*, *156*, 279–296.

Van Beuningen, C., De Jong, N., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, *62*(1), 1–41.

Van Lier, L. (1991). Inside the classroom: Learning processes and teaching procedures. *Applied Language Learning*, *2*, 29–68.

Van Patten, B. (1990). Attending to form and content in the input. *Studies in Second Language Acquisition*, *12*, 287–301.

Van Patten, B. (1994). Evaluating the role of consciousness in second language acquisition: Terms, linguistic features & research methodology. *AILA Review*, *11*, 27–36.

Van Patten, B. (1996). *Input processing and grammar instruction*. New York: Ablex.

Van Weijen, D., Van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2009). L1 use during L2 writing: An empirical study of a complex phenomenon. *Journal of Second Language Writing, 18,* 235–250.

Watanabe, Y., & Swain, M. (2007). Effect of proficiency differences and patterns of pair interaction on second language learning: Collaborative dialogue between adult ESL learners. *Language Teaching Research, 11*(2), 121–142.

Wigglesworth, G., & Storch, N. (2012). What role for collaboration in writing and writing feedback. *Journal of Second Language Writing, 21,* 364–394.

Williams, J. (2001). Learner-generated attention to form. In R. Ellis (Ed.), *Form focused instruction and second language learning* (pp. 303–346). Malden, MA: Blackwell.

Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing, 21,* 321–331.

Yang, L., & Zhang, L. (2010). Exploring the role of reformulations and a model text in EFL students' writing performance. *Language Teaching Research, 14*(4), 464–484.

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly, 19*(1), 79–101.

Appendices

Appendix A: Grammatical Errors in Learners' Essays

Structure	Example of learner's error	Grammatical type
1. Unnecessary definite article (UDA)	Nowadays with the technology and the speed of the life in …	Syntactical
2. Omission of subject	Everything is about consumerism, so ↓ always is well seen to… … with their possibilities ↓ try to show us what they are	Syntactical
3. Verb + preposition/no preposition	Why people invest money in (ON) this? To show to the world we ...	Syntactical
4. Subject + verb agreement	People is used to seeing ...	Syntactical
5. Other 'Subject + constituent´ agreement	People have problems about our/their existence	Syntactical
Subject + Poss. Adjective agreement	In real life things are not as easy as we see it/them on TV	
Subject + Object Pronoun agreement	Most people have psychological problems with ourselves/themselves	
Subject + Reflexive Pronoun agreement		
6. Comparatives	to be more younger People are the most happy/the happiest when ...	Syntactical
7. Wrong word	Much money usually means they have/are successful.	
8. Preposition	Maybe beauty is nearest of this since body painting until tattoos and piercings	Syntactical
9. Word order	... how work TV programs …	Syntactical
10. Addition of constituent/unnecessary element(s)	When we think about happiness mostly we immediately think about money or beauty	Syntactical
11. Omission of constituent/ necessary element(s)	The laugh of a baby or telling your mother you love her are examples of these simple things	Syntactical
12. Passive voice	Models of life that they show to us/we are shown	Syntactical
13. Question formation	What does it mean beauty for you? / What does beauty mean to you?	Syntactical
14. Expressing purpose	Nowadays most people study for having/to have money	Syntactical
15. Tenses	What happened (s) to people that cannot buy these things? (S. Past vs. S. Present for generalizations)	Syntactical
16. Inflections	How to be succeed/successful	Morphological

	Structure	Example of learner's error	Grammatical type
		Famous brands need promote/ promotion of their products	
17.	Demonstrative adjectives	One of this ideas that have …	Morphological
18.	Possessive (´s)	Play with people emotions	Morphological
19.	Third Person Singular	Fashion always try …	Morphological
20.	Verb Form	As if handsome people would made the idea better They will look depress	Morphological
21.	Modal Verbs	I must to brush my teeth now	Morphological
22.	Pluralisation of adjectives	Perfects bodies	Morphological
23.	Countable vs. uncountable nouns/ Quantifiers	The problem is that for many /a long time	Morphological
24.	Singular for Plural (or vice versa)	People want to be like *a TV star/ TV stars* Travel to another *country/ other countries*	Morphological
25.	Irregular Plurals	Instruments in our lifes	
26.	Gerunds (Verb + infinitive/gerund, Gerund as infinitive, Gerund after preposition, Gerund as noun)	The importance of have a 'good look' We think that to look young is Industries try selling us … Another thing the media does is to sell a lot of things	Morphological

Appendix B: Non-grammatical Errors in Learners' Essays

Type of Error	Example of learner's error
Spelling	The *thrut/thruth* is… / *phisicology/Psychology*…
Lexis: Choice of incorrect word or expression	*Actually/Nowadays* the mass media impose the idea of If you *use/wear* the clothes, the handsome man *uses/wears*…
Punctuation & capitalization	It is in difficult moments when people see the truth about themselves in most cases they are afraid of this because... *It is in difficult moments when people see the truth about themselves. In most cases, they are afraid of this because...*
Pragmatic: Appropriate language for formal writing; contractions are not accepted	It's the same around the world People aren't soldiers
Cohesive devices: Omission, incorrect or unclear use of cohesion and coherence devices (referents, connectors). Omission of constituents for inter-sentential clarity.	People do not have their own idea of success, they just follow the idea they watch on TV and/however, if they do not have this they are unhappy
Attention to connectors especially when there was overuse of 'but' and 'and'	They (unclear referent)/Advertisers play with our affective needs

Appendix C: Error Frequency Analysis of Essay 1 (4 groups)

Type of error & Error category		DIR (1-15)	REF (16-30)	SELF (31-45)	CON (46-60)	Totals	%
NON-GRAMMATICAL							
Spelling		124	103	142	98	467	15.5
Lexis		122	86	98	65	371	12.3
Punct. & Capit.		49	55	46	33	183	6.1
Pragmatic		15	10	10	13	48	1.6
Cohesive Devices		17	24	9	7	57	1.9
GRAMMATICAL							
Syntactical	**Structure**						
	UDA	40	40	43	25	148	4.9
	Omission of Subject	8	10	8	2	28	0.9
	Verb + Prep/No Prep	22	12	14	14	62	2.1
	Subj. + Verb Agr.	23	15	15	9	62	2.1
	Subj.+ Other Const Agr.	8	7	7	10	32	1.1
	Comparatives	8	5	4	2	19	0.6
	Tenses	22	17	18	20	77	2.6
	Wrong Word	31	49	89	71	240	8.0
	Preposition	34	30	33	23	120	4.0
	Word Order	23	29	33	36	121	4.0
	Addition of Constituent	23	53	73	59	208	6.9
	Omission of Constituent	38	53	85	63	239	7.9
	Reorganization of Sentence/Phrase	19	37	8	0	64	2.1
	Passive Voice	2	3	3	1	9	0.3
	Question Formation	3	1	2	1	7	0.2
	Expressing Purpose	4	0	0	0	4	0.1
Morphology	**Structure**						
	Inflections	21	13	16	10	60	2.0
	Demonstratives	11	3	12	8	34	1.1
	Possessive's	5	3	9	3	20	0.7
	Third Pers. Singular	15	21	12	15	63	2.1
	Verb Form	18	22	17	12	69	2.3
	Modal Verbs	3	4	11	1	19	0.6
	Pluralization Adject..	4	3	2	3	12	0.4
	Count.-Uncount/ Quantifiers	1	0	2	0	3	0.1
	Singular for Plural	8	15	24	24	71	2.4
	Irregular Plurals	5	4	2	8	19	0.6
	Gerunds	15	22	17	20	74	2.5
		741	749	864	656	3010	
	Totals	3010				3010	

Appendix D: Error Frequency Grid Comprehensive EC

Writing stage: _____ Student ID: _____ Date: _____

Instructions:

Grammatical errors are those that break the rules that govern the structure of words (morphology) and sentences (syntax). In grammar, a constituent is a linguistic unit. Constituency is the relationship between a constituent and the larger unit that it is a part of. A constituent can be a *morpheme, word, phrase, or clause*. For instance, all the words and phrases that make up a clause are said to be *constituents* of that clause.

Non-grammatical errors include semantics (lexis or vocabulary), spelling, punctuation and capitalization, pragmatics (appropriate language for formal writing; contractions are not be accepted), cohesive devices (connectors and ambiguous/unclear or incorrect referents).

Total no. of errors:	
Non Grammatical Errors:	**Grammatical Errors:**
	Sentences (Syntax)
Spelling	Unnecessary definite article (UDA)
Lexis	Omission of subject
Punctuation & Capitalization	Verb + preposition/no preposition
Pragmatic Errors	Subject + Verb agreement
Cohesive Devices	Other 'subject + constituent' agreements
	Comparatives
	Tenses
	Wrong Word
	Preposition
	Word order
	Addition of constituent/non-necessary element(s)
	Omission of constituent/necessary element(s)
	Reorganization of sentence or phrase
	Passive Voice
	Question formation
	Expressing purpose
	Words (morphology)
	Inflections
	Demonstrative Adjectives
	Possessive's
	Third Person Singular
	Verb Form
	Modal Verbs
	Pluralization of adjectives
	Countable-Uncountable + Quantifiers
	Singular for Plural (vice versa)
	Irregular Plurals
	Gerunds

Appendix E: Error Frequency Grid Semi-Comprehensive EC

Selected Errors

Writing stage: _____ Student ID: _____ Date: _____

Overall Accuracy:	
Non-grammatical:	**Grammatical:**
Spelling:	**Syntax:** Omission of Constituent: UDA: Subject + Verb Agreement:
Lexis:	**Morphology:** Third Person Singular: Gerunds:

Appendix F: Noticing Sheet 1

Universidad Nacional Autónoma de México
Centro de Enseñanza de Lenguas Extranjeras

NOTICING SHEET 1 – COMPOSING

Student ID: _____ Group: _____ Date: _____

INSTRUCTIONS:

1. Write down any problems you had with the use of English while writing this essay. Include everything you consider a problem and **specify it,** e.g. if your problem was the use of a particular preposition: do not say 'prepositions' but 'the preposition that corresponds to *worry*' for instance.
2. You may use Spanish to explain your problem and give examples.
3. Leave the sheet blank if you want to.

Appendix G: Noticing Sheet 2

Universidad Nacional Autónoma de México
Centro de Enseñanza de Lenguas Extranjeras

NOTICING SHEET 2 – FEEDBACK

Student ID: _____ Group: _____ Date: _____

INSTRUCTIONS

Direct Error Correction	**Reformulation**	**Self-Correction**
Look at the corrected version of your essay and **compare it with** your original text. 1. Select at least **FIVE DIFFERENT** things (choose more if you want to) that call your attention. Write them on this sheet. 2. Is there anything in your corrected text that particularly attracts your attention? Is there anything that stands out as interesting or strange? If so, what? Leave this sheet blank if you want to.	Look at the reformulated version of your essay and **compare it with** your original text. 1. Select at least **FIVE DIFFERENT** things (choose more if you want to) that call your attention. Write them on this sheet. 2. Is there anything in your reformulated text that particularly attracts your attention? Is there anything that stands out as interesting or strange? If so, what? Leave this sheet blank if you want to.	Read the original version of your essay, **identify** *errors/ bits of language* you can *correct/ improve* yourself and **correct/ improve** them. Write each mistake and the self-corrected version on this sheet. 1. Select **at least FIVE** of the things (choose more if you want to) you *corrected/ improved* that call your attention. Write them on this sheet. 2. Is there anything in your self-correction that particularly attracts your attention? Is there anything that stands out as interesting or strange? If so, what?

Appendix H: Writing Task 1 – Composing

Universidad Nacional Autónoma de México
Centro de Enseñanza de Lenguas Extranjeras

WRITING TASK 1 – COMPOSING

Student ID: _____ Group: _____ Date: _____

INSTRUCTIONS:

1. You have one hour to write an opinion essay, 300 words maximum, on the topic below.
2. Type your essay directly on the PC. Spelling and orthography functions have been deactivated.
3. Your essay will be assessed on its clarity and effectiveness in expressing your ideas, as well as on topic development, organization, accuracy and the appropriateness of grammar and vocabulary.
4. Formal language is required, contractions are not accepted.
5. If you change the writing topic, your essay will be void.
6. Please remember to give your essay a suitable title.
7. Revise your essay before you hand it in.
8. Include the word count at the end of the essay. Do not include the title.

Mass media often impose ideas of what it means to be beautiful, making many young men and women worry about their physical appearance. Should we pay attention to these ideas? Are they important or not? Choose one side and write an essay supporting your opinion. Give reasons for your answer and include examples from your knowledge or experience.

Appendix I: Writing Task 2 – Rewriting

Universidad Nacional Autónoma de México
Centro de Enseñanza de Lenguas Extranjeras

WRITING TASK 2 – REWRITING

Student ID: _____ Group: _____ Date: _____

INSTRUCTIONS:
After looking carefully at your first written essay, ***rewrite*** the same essay with the same prompt. Instructions are the same as in essay 1. It is not a memory test, you are only being asked to repeat the task, thus **changes are expected.**

1. You have one hour to write an opinion essay, **300 words** maximum, on the topic below.
2. Type your essay directly on the PC. Spelling and orthography functions have been deactivated.
3. Your essay will be assessed on its clarity and effectiveness in expressing your ideas, as well as on topic development, organization, accuracy and the appropriateness of grammar and vocabulary.
4. Formal language is required, contractions are not accepted.
5. If you change the writing topic, your essay will be void.
6. Please remember to give your essay a **suitable title**.
7. **Revise** your essay before you hand it in.
8. Include the **word count** in the end of the essay. Do not include the title.

Mass media often impose ideas of what it means to be beautiful, making many young men and women worry about their physical appearance. Should we pay attention to these ideas? Are they important or not? Choose one side and write an essay supporting your opinion. Give reasons for your answer and include examples from your knowledge or experience.

Appendix J: Writing Task 3 – New Writing

Universidad Nacional Autónoma de México
Centro de Enseñanza de Lenguas Extranjeras

WRITING TASK 3 – NEW WRITING

Student ID: _____ Group: _____ Date: _____

INSTRUCTIONS:

1. You have one hour to write an opinion essay, **300 words** maximum, on the **new topic** below.
2. Type your essay directly on the PC. Spelling and orthography functions have been deactivated.
3. Your essay will be assessed on its clarity and effectiveness in expressing your ideas, as well as on topic development, organization, accuracy and the appropriateness of grammar and vocabulary.
4. Formal language is required, contractions are not accepted.
5. If you change the writing topic, your essay will be void.
6. Please remember to give your essay a **suitable title**.
7. **Revise** your essay before you hand it in.
8. Include the **word count** in the end of the essay. Do not include the title.

Advertising on TV, in the movies, newspapers and magazines and on the Internet influence people's ideas and behaviour about **happiness, success** and **values** in general. Is it important to follow these trends? Write an essay to express your opinion. Give reasons for your answer and include examples from your knowledge or experience.

Appendix K: Exit Questionnaire

Universidad Nacional Autónoma de México
Centro de Enseñanza de Lenguas Extranjeras

EXIT QUESTIONNAIRE

Learners' response to and impressions of error-correction techniques

Student ID: _____ Group: _____ Date: _____

INSTRUCTIONS:
Read the instructions below carefully before answering the questions. There are no correct or incorrect answers. Your answers are anonymous. Use **Spanish** in your answers. Thank you for your time and help.

1. What is your impression (opinion or feeling) about the error correction technique offered to you? Advantages and disadvantages.

2. Did the identification of problems you had while writing (Session 1) predispose/prepare you to pay (more) attention to your feedback (Session 2)? Why?

3. What were your criteria to select FIVE things to attend to (NOTICING SHEET-2)?

4. What would you advise teachers to make error correction efficient for you?

Student's profile:

Date of birth: _____ Major: _____ Years of formal English study: _____

☺ ***Thank you!***

Appendix L: Informed Consent Documentation

From: Ethics (RSO) Enquiries
Sent: 28 April 2014 15:22
To: Solares-Altamirano, Maria-Elena
Subject: Ethics approval

Dear Maria

Thank you for submitting your completed stage 1 self-assessment form for The effects of different noticing conditions during WCF on EFL learners' writing skill. I can confirm that approval has been granted for this project.

As principal investigator your responsibilities include:

- ensuring that (where applicable) all the necessary legal and regulatory requirements in order to conduct the research are met, and the necessary licenses and approvals have been obtained;
- reporting any ethics-related issues that occur during the course of the research or arising from the research (e.g. unforeseen ethical issues, complaints about the conduct of the research, adverse reactions such as extreme distress) to the Research Ethics Officer;
- submitting details of proposed substantive amendments to the protocol to the Research Ethics Officer for approval. Please contact the Research Ethics Officer, Debbie Knight (ethics@lancaster.ac.uk 01542 592605) if you have any queries or require further information.

Kind regards,

Debbie

Debbie Knight
Research Ethics Officer, Research Support Office
B58, B Floor, Bowland Main
Lancaster University. Lancaster, LA1 4YT
Email: ethics@lancaster.ac.uk
Tel 01524 592605
Web: Ethical Research at Lancaster:
http://www.lancs.ac.uk/depts/research/lancaster/ethics.html

December 13th, 2013.

Subject: **Letter of Authorization**

To whom it may concern,

This letter serves as authorization to Maria Elena Solares Altamirano to carry out her PhD research at the Foreign Language Teaching Centre of the National University of Mexico (CELE, UNAM). We know Maria Elena is currently registered in the PhD programme in Applied Linguistics at Lancaster University and we will be happy to have her perform her research in our language centre.

Maria Elena has provided us with the Information Sheet that explains the objectives of her study and has explained to us further details. Consent Forms for the students and teachers she will require have also been provided. We believe engaging learners and teachers from our centre will be of great benefit for our institution. Therefore, we will give Maria Elena all the support and coordination necessary to best help her perform her study.

Sincerely yours

_____

Dra. Aine Signoret
Director of the Foreign Language Teaching Centre
CELE, UNAM

c.c.p. Lic. Joaquín Martínez.- Head of the English Department
c.c.p. Lic. Bertha López Escudero.- Secretaria General

October 2014

STUDENT'S INFORMATION SHEET

As part of my Doctoral studies in the Department of Linguistics and English Language, I am carrying out a study about EFL (English as a Foreign Language) learners' response to different error correction conditions in a writing task. My study involves students' participation in a four-stage writing task and answering a questionnaire.

I have approached you because I am interested in -the writing performance and response to different error correction conditions- of upper-intermediate learners studying English at CELE, UNAM. I would be extremely grateful if you agree to take part in my study.

If you decide to participate, this will involve the following: (1) you will participate in a four-stage writing task, each stage on a different day; (2) you will reflect on the feedback given to your writing or on the ECC offered to you; (3) you will answer a questionnaire at the end of the study. These activities will take place during your class schedule. But in a different classroom to give you access to a computer.

You are free to withdraw from the study at any time. If you withdraw while the study takes place or until one month after the study finishes, I will not use any of the information that you provided. If you withdraw later, the information you shared with me will be used as part of the study. At every stage, your name will remain anonymous. The information you provide (a) will be kept securely in locked drawers and encrypted documents and (b) will be used for academic purposes only. This will include my PhD thesis, conference presentations, future teaching contexts, journal articles and other academic publications.

If you have any queries about the study, please feel free to contact myself or my supervisor, Dr. Patrick Rebuschat who can be contacted at p.rebuschat@lancaster.ac.uk or by phone on +44 1524 592433 or on +44 1524 592434. You may also contact the Head of Department, Prof. Elena Semino, on +44 (0)1524 594176.

Signed

_____

Maria Elena Solares
m.solares-altamirano@lancaster.ac.uk
Mobile: 04455-3136-2981

Lancaster University
Lancaster LA1 4YL
United Kingdom
Tel: +44 (0)1524 593045
Fax: +44 (0)1524 843085
http://www.ling.lancs.ac.uk

UNIVERSITY OF LANCASTER

Department of Linguistics and English Language

Student's Consent Form

Project title:

Learner-initiated noticing in three comprehensive error correction conditions: its effects on learners' writing accuracy

1. I have read and had explained to me by Maria Elena Solares the Information Sheet relating to this project.
2. I have had explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements described in the Information Sheet in so far as they relate to my participation.
3. I understand that my participation is entirely voluntary and that I have the right to withdraw from the project any time, but no longer than one month after its completion. If I withdraw after this period, the information I have provided will be used for the project.
4. I have received a copy of this Consent Form and of the accompanying Information Sheet.

Name: _____

Signed: _____

Date: _____

List of Appendices

List of Tables

List of Figures